Juho Liski

Adaptive Hear-Through Headset

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 25.1.2016

Thesis supervisor and advisor:

Prof. Vesa Välimäki

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Juho Liski

Title: Adaptive Hear-Through Headset

Date: 25.1.2016 Language: English Number of pages: 9+69

Department of Signal Processing and Acoustics

Professorship: Audio Signal Processing

Supervisor and advisor: Prof. Vesa Välimäki

Hear-through equalization can be used to make a headset acoustically transparent, i.e. to produce sound perception that is similar to perception without the headset. The headset must have microphones outside the earpieces to capture the ambient sounds, which is then reproduced with the headset transducers after the equalization. The reproduced signal is called the hear-through signal. Equalization is needed, since the headset affects the acoustics of the outer ear.

In addition to the external microphones, the headset used in this study has additional internal microphones. Together these microphones can be used to estimate the attenuation of the headset online and to detect poor fit. Since the poor fit causes leaks and decreased attenuation, the combined effect of the leaked sound and the hear-through signal changes, when compared to proper fit situation. Therefore, the isolation estimate is used to control the hear-through equalization in order to produce better acoustical transparency. Furthermore, the proposed adaptive hear-through algorithm includes manual controls for the equalizers and the volume of the hear-through signal.

The proposed algorithm is found to transform the used headset acoustically transparent. The equalization controls improve the performance of the headset, when the fit is poor or when the volume of the hear-through signal is adjusted, by reducing the comb-filtering effect due to the summation of the leaked sound and the hear-through signal inside the ear canal. The behavior of the proposed algorithm can be demonstrated with an implemented Matlab simulator.

Keywords: Acoustic measurements, acoustic signal processing, augmented reality, digital filters, headphones

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Juho Liski

Työn nimi: Adaptiiviset läpikuuluvuuskuulokkeet

Päivämäärä: 25.1.2016 Kieli: Englanti Sivumäärä: 9+69

Signaalinkäsittelyn ja akustiikan laitos

Professuuri: Audiosignaalinkäsittely

Työn valvoja ja ohjaaja: Prof. Vesa Välimäki

Läpikuuluvuusekvalisoinnilla voidaan saavuttaa akustinen läpinäkyvyys kuulokkeita käytettäessä, eli tuottaa samankaltainen ääniaistimus kuin mikä havaittaisiin ilman kuulokkeita. Käytetyissä kuulokkeissa tulee olla mikrofonit kuulokkeen ulkopinnalla, joiden avulla voidaan tallentaa ympäröiviä ääniä. Mikrofonisignaalit ekvalisoidaan, jolloin niistä tulee läpikuuluvuussignaalit, ja toistetaan kuulokkeista. Ekvalisointi on tarpeellista, sillä kuulokkeet muuttavat ulkokorvan akustiikka ja siten myös äänihavaintoa.

Tässä diplomityössä käytetyssä prototyyppikuulokeparissa on edellä mainittujen mikrofonien lisäksi myös toiset, korvakäytävän sisälle asettuvat mikrofonit. Yhdessä näiden kahden mikrofonin avulla voidaan määrittää reaaliaikainen estimaatti kuulokkeen vaimennukselle ja tunnistaa huono istuvuus. Koska huonosti asetettu kuuloke vuotaa enemmän ääntä korvakäytävän sisään kuin kunnolla asetettu, kuulokkeen äänen ja vuotavan äänen yhteisvaikutus muuttuu. Tästä syystä vaimennusestimaattia käytetään läpikuuluvuusekvalisoinnin säätöön, jotta akustinen läpinäkyvyys ei kärsisi. Lisäksi esitellyssä algoritmissa on manuaaliset säädöt ekvalisaattoreille ja läpikuuluvuussignaalin voimakkuudelle.

Esitetyn algoritmin havaitaan tuottavan akustinen läpinäkyvyys, kun sitä käytetään prototyyppikuulokkeiden kanssa. Ekvalisointisäädöt parantavat kuulokkeiden toimintaa istuvuuden ollessa huono tai säädettäessä läpikuuluvuussignaalin voimakkuutta, koska ne vähentävät kampasuodatusefektiä, joka voi aiheutua vuotavan äänen ja läpikuuluvuussignaalin summautuessa. Esitellyn algoritmin toimintaa voidaan havainnollistaa toteutetulla Matlab-simulaattorilla.

Avainsanat: Akustinen signaalinkäsittely, akustiset mittaukset, digitaaliset suotimet, lisätty todellisuus, kuulokkeet

Preface

This thesis has been done for Department of Signal Processing and Acoustics with funding from Nokia Technologies.

I want to thank my supervisor Professor Vesa Välimäki for the excellent guidance, the people from Nokia Technologies who helped me with the project, my co-workers in the department, my family and friends for the support during my studies, and, finally, my girlfriend for all the love and support.

Otaniemi, 25.1.2016

Juho Liski

Contents

Symbols and abbreviations

Symbols

a	frequency parameter of the Regalia-Mitra filter
a_k	predictor coefficients
$A_{inv}(z)$	all-zero inverse filter
$A(z)$	transfer function of an allpass filter
B	bandwidth of digital resonator
BW	bandwidth of the Regalia-Mitra filter
c	speed of sound
$d(k)$	desired signal
$D(\omega)$	group delay in seconds
$e(k)$	error signal
$e_p(k)$	*a posteriori error* signal
f	frequency in Hertz
f_c	center frequency
$f_{r,closed}$	first half-wavelength resonance of a closed ear canal
$f_{r,open}$	first quarter-wavelength resonance of an open ear canal
f_s	sampling frequency
f_{tp}	turning point frequency
g_d	gain of direct sound
g_{eq}	gain of pseudoacoustic representation
G	gain of the Regalia-Mitra filter in dB
$h(t)$	impulse response of a system
$H(z)$	transfer function of a system
K	gain of the Regalia-Mitra filter
l	physical length of the ear canal
l_{eff}	effective length of the ear canal
L	length of LMS algorithm
L_D	delay in samples
M	filter order
N	LP order
p	real pole
$P(\omega)$	phase delay in seconds
Q	quality factor of a filter
r	radius of the ear canal
r_k	autocorrelation coefficients
R	pole radius
$s(n)$	original signal
$\hat{s}(n)$	predicted signal
T	length of sine sweep in seconds
$w(k)$	warped signal or impulse response
$\mathbf{w}(k)$	LMS filter coefficient vector
$x(t)$	input signal (sine sweep)
$y(t)$	output signal

α	memory factor
$\delta(k)$	positive sequence
Δf_{ERB}	ERB bandwidth
θ	pole angle
$\Theta(\omega)$	phase response
κ	signal color parameter
λ	allpass filter parameter
μ	step size
$\nu(k)$	noise consisting of background measurement noise and impulse noise
ψ	center frequency of digital resonator
ω_0	normalized center frequency of the Regalia-Mitra filter
ω_1	initial radian frequency of a sine sweep
ω_2	final radian frequency of a sine sweep
Ω_1	normalized filter cutoff frequency of the Regalia-Mitra filter
Ω_2	normalized filter bandwidth of the Regalia-Mitra filter

Abbreviations

AD	analog-to-digital
ARA	augmented reality audio
DA	digital-to-analog
DC	direct current
DRP	drum reference point
DSP	digital signal processing
ERB	equivalent rectangular bandwidth
FFT	fast Fourier transform
FIR	finite impulse response
GPS	Global Positioning System
IFFT	inverse fast Fourier transform
IHL	in-head localization
IIR	infinite impulse response
ILD	interaural level difference
ITD	interaural time difference
ITU-T	International Telecommunication Union, Telecommunication Standardization Sector
LMS	least mean square algorithm
LP	linear prediction
MSE	mean square error
NLMS	normalized least mean square algorithm
NSA	normalized sign algorithm
RVSS-NLMS	robust variable step size normalized least mean square algorithm
SPL	sound pressure level

1 Introduction

The use of headphones is constantly increasing due to digital music and smartphones: more and more people have equipment to use headphones whenever they want, since every smartphone includes a hands-free headset, i.e. headphones with an attached microphone. According to Gartner Inc., the global smartphone sales amounted to 1.2 billion units in 2014 [1], and the first quarter of 2015 showed further increase in sales when compared to the same quarter of 2014 [2].

As a consequence, headphones are nowadays used in many different environments. These include the stationary home and office settings, but especially mobile settings, such as commuting and travelling. This results in increased exposure to ambient sounds while wearing the headphones. However, in addition to unwanted noise, these ambient sounds can also be desirable, such as speech, or even vital, such as sirens of emergency vehicles. Therefore, the attenuation properties of headphones may be viewed as nuisance, since the headphones must be removed to properly hear the environmental sounds. A more convenient solution, however, would eliminate the necessity to remove the headphones, and this is exactly the ideas behind augmented reality audio (ARA) and hear-through technology.

Augmented reality audio applications are based on an idea of constantly wearing headphones. The premise for this is a hear-through signal, which is a processed version of the ambient noise that is simultaneously reproduced with the headphone. The hear-through signal is supposed to transform the headphone acoustically transparent, i.e. produce similar perception of sound that would be heard without the headphones. After that, additional augmented sounds may be reproduced with the headphones to further enhance the perceived sound environment.

Acoustical transparency is desirable, since people are used to the response of their ears while perceiving sounds. However, headphones alter the acoustics of the outer ear, which results in an altered listening experience. Thus, the hear-through signal can be used to better include headphones to everyday life. Furthermore, the increased number of different virtual reality and augmented reality glasses demands good understanding of ARA and hear-through functions, since the visual aspects can be supported with audio and vice versa. Finally, ARA and hear-through functions could prove especially advantageous in working life. For example, with the accelerating globalization the communication between different branches of a company may become difficult. Teleconferences via phone may be improved, if the participants are virtually seated around a table while wearing headphones so that both the actual and virtual persons are easily distinguished from each other with the help of ARA instead of reproducing the speech of every virtual participants with a single loudspeaker.

The purpose of this thesis work is to develop an algorithm with two functions: firstly it should constantly estimate the isolation of a prototype headset, and secondly the effects of the headset on the acoustics of the ear should be compensated. The isolation estimation should be performed online, and instead of controlled signals, ambient sounds should be used. The estimated isolation can then be used, for example, to inform the user of a lacking fit of the headset. However, the main use for the isolation estimate is to control the hear-through equalization. The default target

response of the hear-through function is the aforementioned acoustical transparency. However, since the fit of the headphones affects both its isolation and frequency response, the equalization is adaptive in order to provide acoustical transparency even with poor fits. Furthermore, the user should also have controls over the hear-through equalization and the overall volume of the hear-through signal to possibly improve it for certain situations.

In order to accomplish the targets of this thesis work, the following main objectives are outlined:

1. define a way to estimate the isolation using the internal and external microphones of the prototype headset
2. measure the characteristics of the prototype headset and produce a hear-through equalization
3. define an automatic control for the equalization according to the isolation estimation

When these objectives are accomplished, the algorithm should be able to continuously adapt an estimate of attenuation and use it to fine-tune acoustical transparency.

The structure of the thesis is organized as follows. Firstly, the theory is described in two parts: Section 2 recapitulates headphones and ARA theory, while Section 3 covers digital filters. In the former, the acoustics of the headphones are reviewed as well as headphone measurements, and thus it helps to tackle the second objective. On the other hand, in the latter theory section adaptive filters and equalizers are discussed, which are utilized to accomplish the first and third objective. After the theory sections, the proposed algorithm is presented and its building blocks are analyzed in Section 4. In the following Section 5, measurement results are presented, which validate the behavior of the proposed algorithm. Finally, the thesis work is summarized in Section 6.

2 Headphones and Augmented Reality Audio

In this section basic headphone theory is discussed as well as some measurement techniques for headphones. The main focus is on in-ear headphones, because they are most commonly used with mobile devices and are thus used in this thesis work. The measurement techniques presented here are related to the headphone frequency response and ambient isolation measurements, since they are the most representative attributes of headphones in relation to the sound quality in different environments. After headphones, augmented reality audio (ARA) is presented together with hear-through systems. The basic theory of ARA is essential to this work, and the ARA applications act as motivation for the whole thesis.

The structure of the section is as follows: First the different types of headphones are presented as well as different types of traducers used in headphones. After that the ambient sound isolation of headphones is discussed. The next subsection recapitulates in-ear headphone acoustics including the pressure chamber principle and the occlusion effect. Next, a measurement technique for headphones is presented. Finally, ARA theory is presented as well as ARA applications and the comb-filtering effect, which can be a significant problem in ARA systems.

2.1 Headphone Types

ITU Telecommunication Standardization Sector (ITU-T) distinguishes four different types of headphones, which differ in size, appearance and intended application [3]:

1. circum-aural
2. supra-aural
3. intra-concha
4. insert/in-ear headphones.

These different types of headphones are shown in Fig. 1. In Fig. 1(a) a circum-aural headphone is pictured. It covers the whole ear and sits against the head. In comparison, a supra-aural headphone only covers the pinna while sitting on top of the ear [4, 5], like in Fig. 1(b). Figure 1(c) shows an intra-concha headphone, which is much smaller than the previous types and is placed loosely in the concha cavity. Finally, an insert or in-ear headphone [4, 6] is shown in Fig. 1(d). It resembles an earplug, because it is inserted tightly into the ear canal blocking it completely. For example the desired sound quality or the needed ambient isolation can dictate which headphone type is suitable for each application or listening environment.

When one or more microphones are attached to headphones, they become a headset [7]. Often a mono microphone is used in the wire of the headphones. However, if two microphones are used in the earphones itself so that the microphones are located near the ear canal entrances, binaural signals can be recorded [8]. This means that the directional cues from the person's head and body are preserved and can be played back later with headphones.

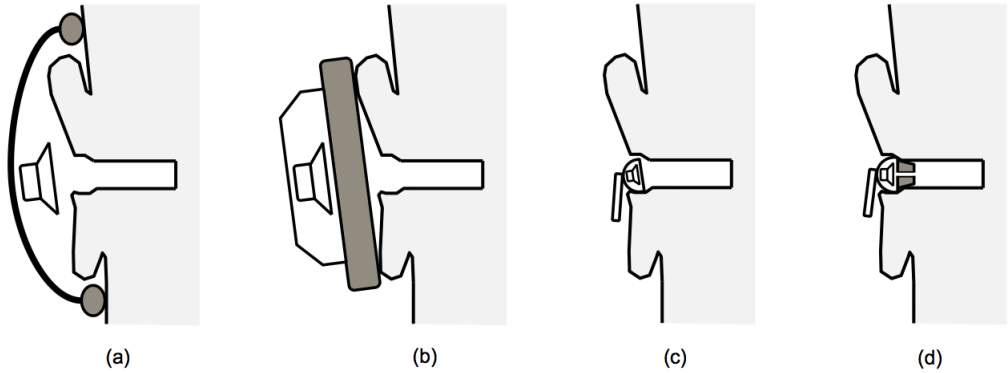

Figure 1: Different types of headphones according to ITU-T Recommendation P.57: (a) circum-aural (b) supra-aural (c) intra-concha (d) in-ear/insert headphone. From [9].

2.1.1 Transducer Types

In addition to the size and appearance, headphones can also be categorized by the used transducer type. One common type is the moving-coil, or dynamic transducer [4, 10]. It resembles a small loudspeaker in operation, because the diaphragm acts similarly to the cone of loudspeakers. The diaphragm is attached to the voice coil, i.e. a coil of wire, which is situated in a magnetic field. When the music signal is driven through the voice coil, the magnetic field causes the voice coil and also the diaphragm to move and produce sound.

Another type of transducer is the isodynamic transducer [4]. In contrast to the dynamic transducer, in the isodynamic one the conducting wire is attached directly to the diaphragm and the magnetic field is produced around it. Thus the whole diaphragm is driven in phase.

Third type are electrostatic and electret transducers [4, 10]. In those, the diaphragm is a charged membrane located between two electrodes. The electric field produced by the electrodes is modulated by the sound signal, which moves the diaphragm. The difference between electrostatic and electret transducers is that the electrostatic one requires a polarization voltage from an outside source, whereas in the electret transducer the membrane itself is permanently charged [4]. The auditory presentation of electrostatic transducers is regarded as clear, which is the result of small moving masses not causing severe phase distortions [5].

Finally, an electromagnetic transducer, or a balanced armature, is a miniature transducer, which was originally used in telephone receivers, hearing aids and in military applications [11]. Long they were not deemed suitable for high-fidelity applications because of the large mass of vibrating components [4], but nowadays they are also found in high-fidelity in-ear headphones [12]. Balanced armature consists of a coil, pivoted armature, permanent magnet and a diaphragm [13]. The armature

is connected to the diaphragm and balanced between magnets. When a music signal is driven through the coil, the magnetic field changes and causes the armature and, thus, the diaphragm to move.

2.2 Ambient Sound Isolation

Ambient sound isolation, or sound insulation [4], of headphones expresses how much external sounds are attenuated by the headphones when they are worn. Depending on application the ambient isolation can be seen as both pro and con: for example high isolation can mask environmental stimuli and thus cause dangerous situations for pedestrians in cities [14], but on the other hand in an ARA hear-through system high isolation is beneficial, because it can diminish the comb filtering effect [15].

The isolation of headphones is determined by both the structure of the headphones and the coupling of the headphones to the head or ear [4]. First, the effects of a structural design choice of supra-aural and circum-aural are presented. Even though these headphone types can be tightly coupled to the ear by sitting around it or on top of it, the isolation varies greatly between open and closed headphones [3, 4, 5]. In closed supra-aural or circum-aural headphones the back of the headphone is solid creating a closed air volume, which attenuates external sounds, whereas open headphones have holes in the back. Alternatively the whole diaphragm of open headphones can be transparent to sound allowing ambient noise to reach the ear almost unattenuated.

In intra-concha and in-ear headphones the main divider in isolation is the coupling to the user's ear. Intra-concha headphones sit normally loosely at the mouth of the ear canal, which results in poor ambient sound isolation. On the other hand, in-ear headphones are placed tightly into the ear canal. Thus, like an earplug, it blocks the ear canal offering much better isolation than intra-concha headphones.

Figure 2 shows some examples of ambient noise isolation of different headphone types. Attenuation of 0 dB means that the headphones do not affect the ambient sound at all, and negative values express the amount of attenuation. As can be seen, isolation is highly dependent on headphone type but also on frequency. As stated, the in-ear headphones function like earplugs and, thus, have the greatest attenuation. This can be seen from the purple curve. The loose fit and resulting poor isolation of intra-concha headphones can be seen from the red curve. In Fig. 2 the difference between closed and open headphones is also seen between the isolation of circum-aural and supra-aural headphones. The circum-aural headphones, illustrated by the black curve, show the poorest isolation of the headphones. This can be explained with the open back that lets the sound through regardless of a possible tight coupling to the head. On the other hand, supra-aural headphones, which sit on top of the pinna, show much better isolation (the green curve), because of the closed back, which prevents sound from reaching the ear.

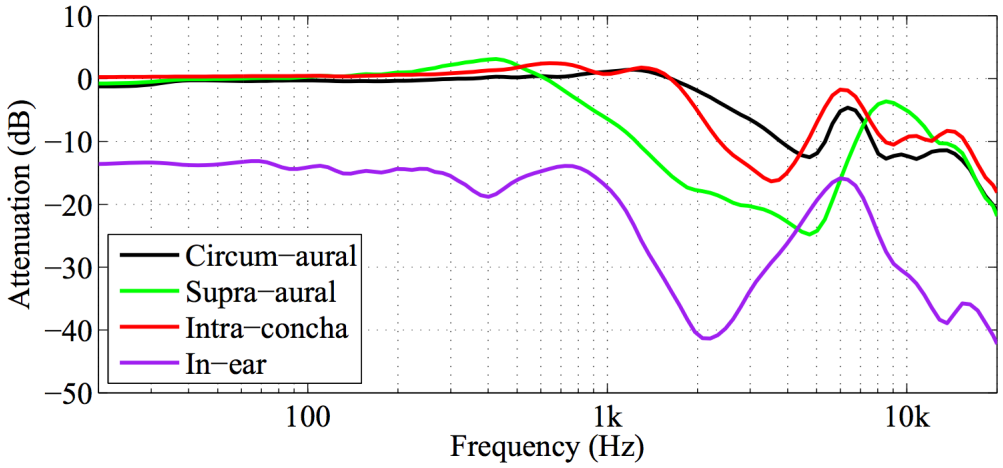

Figure 2: Headphone isolation of different headphone types. From [9].

2.3 In-Ear Headphone Acoustics

Headphone listening differs from normal open ear listening. When listening for example a loudspeaker, the sound field interacts with the environment [16], upper body, head and outer ear of the listener [4] before surrounding the listener's head. With headphones, on the other hand, the sound is played straight to the ear canal and therefore all the aforementioned characteristics are lost [4]. To counter this, international standards offer two different target responses for headphones, free-field and diffuse-field responses. In the former the target response resembles the response from a flat loudspeaker in front of the listener to the ear measured in an anechoic chamber [4]; in the latter measurement is done in a reverberant room and the sound is coming from every direction [17]. Recently, Lorho [18] and Olive *et al.* [19, 20] have suggested their own target curves. Olive's new headphone target response "simulates the in-room response of a high-quality loudspeaker system calibrated in a room," and he found that it was preferred to the previous target curves. These target responses differentiate headphones from other electroacoustic devices, since they usually have perfectly flat target response [4]. Figure 3 shows the aforementioned headphone target responses.

Headphone listening also causes in-head localization (IHL) with normal stereo signals [4]. Since headphones produce high channel separation by playing each channel to the respective ear, interaural time difference (ITD) and interaural level difference (ILD) can not be used to localize the sound [21, 22]. Another factor for IHL is the lack of head movement relative to sound [23]. To decrease the IHL and to even achieve out-of-the-head localization, digital signal processing (DSP) and head trackers can be used. If, for example, a balanced stereo widening network is used [24], cross-talk is added to the stereo signal using DSP, which produces artificial ITD and ILD. When head trackers are utilized, the direction of the music stays the same even if headphone listeners move their heads.

Inserting an in-ear headphone tightly to the ear canal blocks it completely and

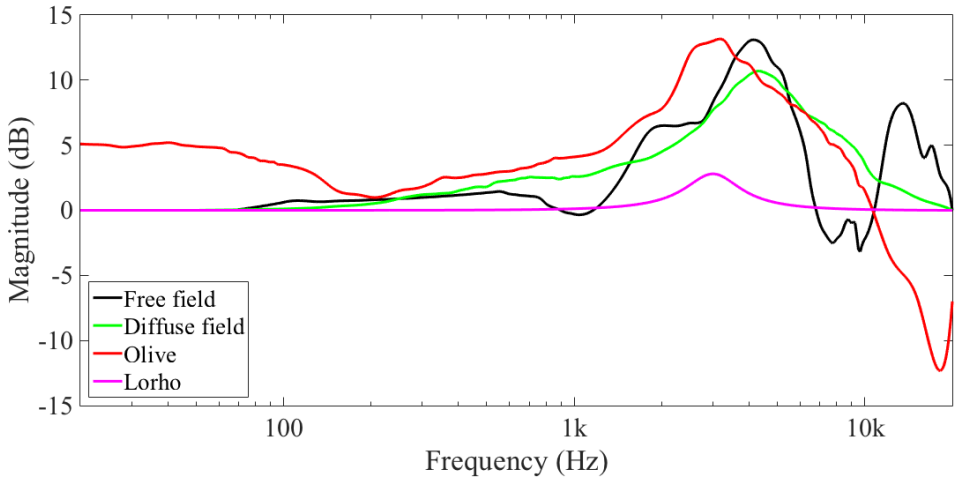

Figure 3: Different headphone target responses: free-field from [25], diffuse-field from [26], Lorho from [18] and Olive from [20].

thus changes the acoustics of the canal [27]. An open ear canal can be approximated as a tube blocked at the other end by the ear drum [28]. When in-ear headphones are used, this tube becomes blocked from both ends. Open ear canal acts as a quarter-wavelength resonator with a first resonance at the frequency [29]:

$$f_{r,open} = \frac{c}{4l_{eff}}, \tag{1}$$

where c is the speed of sound and l_{eff} is the effective length of the ear canal. The effective length of the ear canal is longer than the actual length, since it takes into account the attached-mass of air, and for a flanged, open ended tube it can be calculated as [29, 30]

$$l_{eff} = l + \frac{8r}{3\pi}, \tag{2}$$

where r is the radius and l the physical length of the ear canal.

An average human ear canal is approximately 27 mm in length and 7 mm in diameter [31, 32]. When these values are substituted to Equations (1) and (2) together with the speed of sound of 343 m/s (at 20 °C), the effective ear canal length is 30 mm and the first quarter-wavelength resonance appears at 2860 Hz. According to Wiener and Ross [33] the first resonance of ear canal averages at around 3 kHz and has an amplification of about 20 dB.

However, when in-ear headphones are inserted into the ear-canal, it becomes blocked from both ends and starts acting like a half-wavelength resonator [25]. The first resonance frequency of such system is [29]

$$f_{r,closed} = \frac{c}{2l}. \tag{3}$$

Furthermore, in-ear headphones also shorten the ear canal by approximately 5 mm [9] and raise the temperature inside the ear canal to about 35 °C [34]. This increase

in temperature also increases the speed of sound to 352 m/s. With these new values and Equation (3), the first resonance frequency of a closed ear canal results at 8 kHz. Therefore, in-ear headphones not only cancel a pronounced natural resonance, but also create an unnatural new one.

2.3.1 Pressure Chamber Principle

According to the pressure chamber principle, low and middle frequencies are enhanced when listening to in-ear headphones [4]. When sound is produced into a small cavity, such as to a blocked ear canal, it is easy to produce high SPL levels, because the sound pressure is distributed uniformly in the cavity. This happens, since up to about 2 kHz the wavelength is large in comparison to the dimensions of the cavity [4]. Therefore, inside the ear canal, the pressure is in phase with the volume displacement of the headphone diaphragm and the amplitude of the pressure is proportional to that volume displacement. The pressure chamber principle is most pronounced when there are no leaks present, but still small leaks do not affect it much [4].

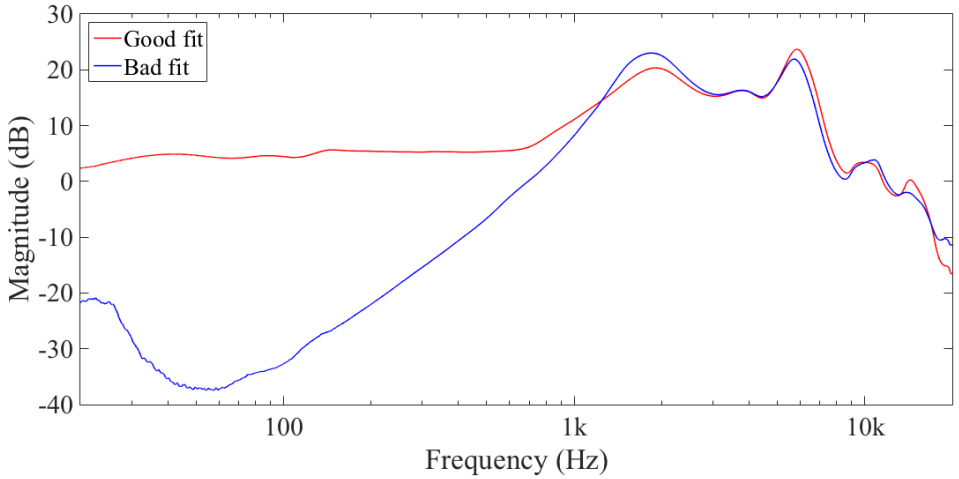

Figure 4: The effect of the pressure chamber principle. In red curve the low frequency response is boosted by the pressure chamber principle, whereas in the blue curve the poor fit results in leaks and lack of boost.

2.3.2 Occlusion Effect

The occlusion effect occurs, when the blocked ear canal results in changed sound pressure and, thus, in changed perception of own voice [35]. Due to the occlusion effect, one's own voice may sound hollow and loud or like talking inside a barrel [36, 37, 38], and therefore it is perceived unnatural [35]. Because the perception of own voice is so vital in normal communication, this can leave the speaker feeling uncomfortable and may even cause problems in communication.

Normally people hear their own voice both via air-conduction and bone-conduction at the same time [39]. These two sources have the same order of magnitude in loudness

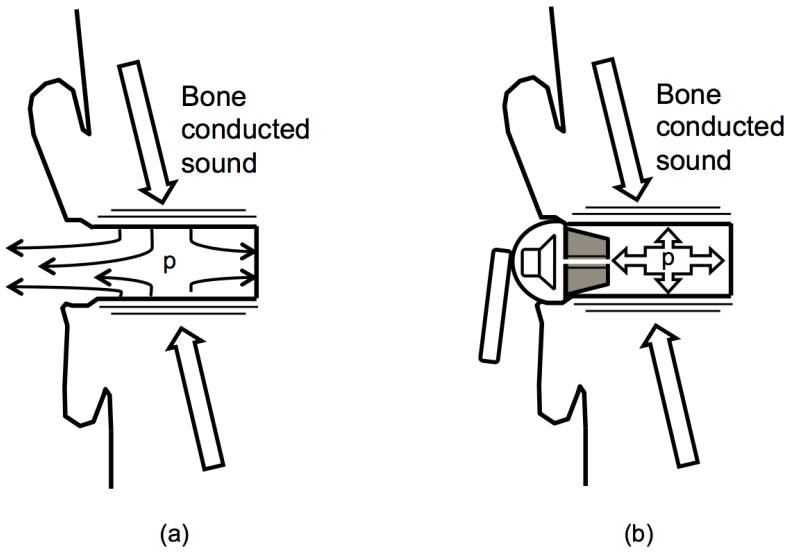

Figure 5: Occlusion effect. In (a) the ear canal is open and the bone-conducted sounds radiate freely to the outside environment. In (b) The ear canal is blocked by an in-ear headphone, which results in the occlusion effect, where the bone-conducted sounds are amplified. From [9].

and they are relevant for perceiving one's own voice naturally [39]. Air-conducted sound comes from the mouth and circles the head to arrive at the ears. The bone-conducted sound, on the other hand, is transmitted through the skull to the ear canal. Nonetheless, when the ear canal is open, like in Fig. 5(a), the bone-conducted sounds can escape from the ear canal to the environment without generating additional sound pressure [4, 36]. However, when the ear canal is blocked with a headphone, like in Fig. 5(b), the bone-conducted sound is trapped inside the ear canal [36, 37, 38]. The trapped sound follows the pressure chamber principle discussed in the previous section, which leads to an amplified perception of the bone-conducted sound [4, 36, 37, 38]. Also, the ambient isolation of the headphones further distorts the perception of own voice, because the air-conducted sound is attenuated at the same time as the bone-conducted sound is amplified [35].

The occlusion effect is especially bothersome in ARA applications and with hearing aids, because people usually talk with them on as opposed to just listening to music with headphones. The effect can be passively reduced by introducing vents to the headset casing [36, 37, 38]. This is a viable option with hearing aids, but with headsets it would result in decreased isolation, which is normally unwanted. Additionally, vents in the casing might cause acoustical feedback, if the headset microphone is situated close to the vents [38].

2.4 Headphone Measurements

In this thesis work two types of headphone measurement were crucial: frequency response and isolation measurements. Both were executed with swept-sine technique. The magnitude of frequency response of headphones describes the headphone's ability to reproduce different frequencies and thus determines the perceived sound quality of the headphones. Measured isolation, as described before, expresses how much ambient sounds are attenuated by the headphones. Noise isolation affects the perceived sound quality, when the headphones are used in noisy environments.

Swept-sine measurements have been discussed in many articles (for example Berkhout *et al.* [40, 41], Clarkson *et al.* [42] and Griesinger [43, 44]), but it became widely used after Farina's article [45], because computational power of computers had increased and software tools became widely available [46]. Farina further discussed some problems of the swept-sine method and solutions to them [46], and both Müller *et al.* [47] and Stan *et al.* have compared the swept-sine technique with other impulse response measurement techniques.

In impulse response measurements a known input signal $x(t)$ is applied to a system. With the swept-sine measurement, this input signal has sinusoidal shape and exponentially increasing frequency, and it can be generated with:

$$x(t) = \sin\left[\frac{T\omega_1}{\ln(\frac{\omega_2}{\omega_1})} e^{\frac{t}{T}\ln(\frac{\omega_2}{\omega_1})} - 1\right], \tag{4}$$

where ω_1 is the initial radian frequency, ω_2 is the final radian frequency, and T is the length of the sweep in seconds [45]. The response $y(t)$ caused by this input signal is then recorded, and the impulse response of the system is obtained with linear deconvolution. First, an inverse filter $f(t)$ is determined that packs the input signal into a Dirac's delta function $\delta(t)$ [45]. Then, deconvolution of the impulse response is acquired by convolving the output signal $y(t)$ with the inverse filter $f(t)$. In Matlab this is realized with the fast Fourier transform (FFT):

$$h(t) = \mathrm{IFFT}\left(\frac{\mathrm{FFT}(y(t))}{\mathrm{FFT}(x(t))}\right), \tag{5}$$

where IFFT is the inverse fast Fourier transform. Figure 6 shows the frequency response of the prototype headset used in this work, and Fig. 7 shows the isolation of the same headset from an angle of 15° from the nose line.

The sine-sweep technique is extremely useful in quiet environments, because it gives both the linear impulse response and the harmonic distortion components [48]. In addition, it yields a high signal-to-noise ratio [46] and it is relatively tolerant against time variance and distortion [47].

2.5 Augmented Reality Audio and Hear-Through Systems

Augmented reality can be defined as a real-time combination of real and virtual worlds [49]. Thus, in ARA a person's natural surrounding sound environment is extended with virtual sounds so that they both are heard together.

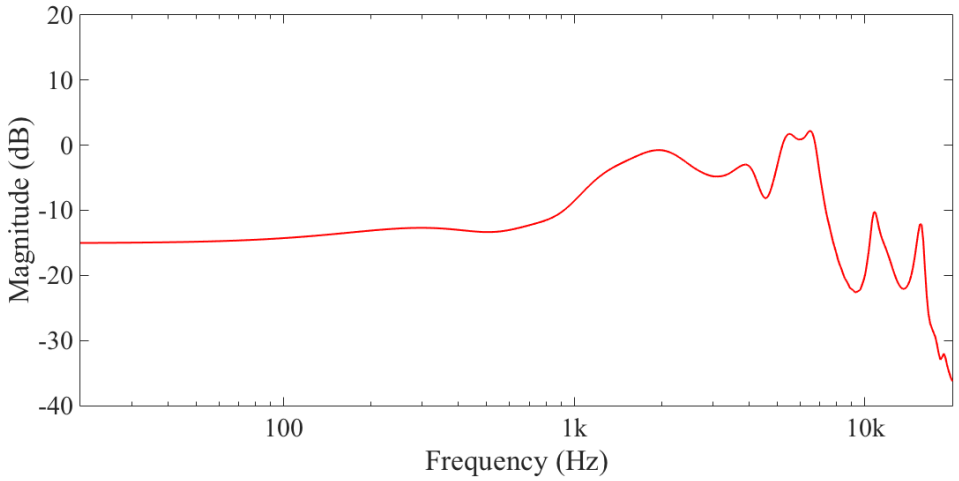

Figure 6: Prototype headset frequency magnitude response.

Figure 8 shows a mobile ARA system. It consists of a headset with binaural microphones and an ARA mixer [7, 50]. In normal usage the microphone signals are routed straight to the headphones, which should produce the surrounding sound environment unaltered to the ears. Therefore, the ARA system should be acoustically transparent [7] and have minimal latency [49]. The copy of the natural sound environment that is captured by the binaural microphones and reproduced with the headset, is called pseudoacoustic environment to separate it from normal listening situation [49]. For this work, the pseudoacoustic representation is the most interesting part of ARA system, since it is basically the hear-through signal, which is the subject of this thesis.

In ARA usage, the sound pressure at the eardrum is a sum of the pseudoacoustic representation and leaked sound, which may lead to comb-filtering effect (see Section

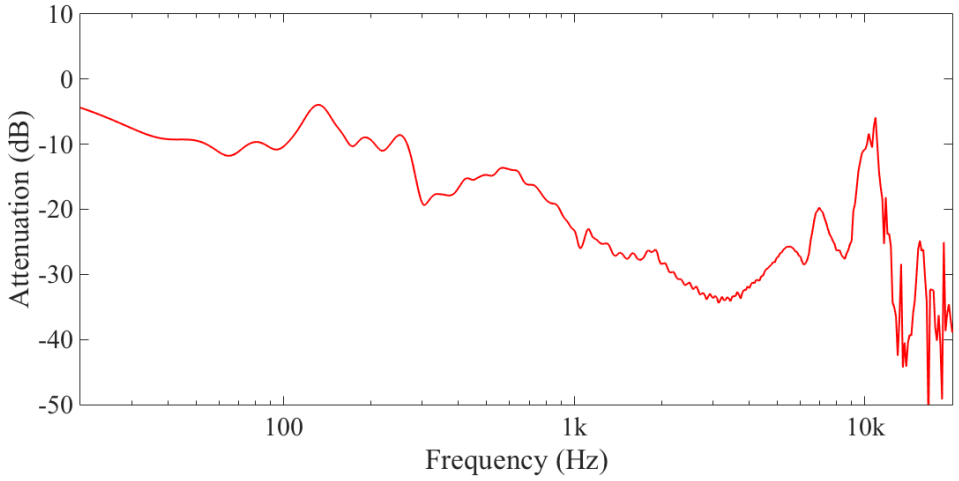

Figure 7: Prototype headset isolation curve.

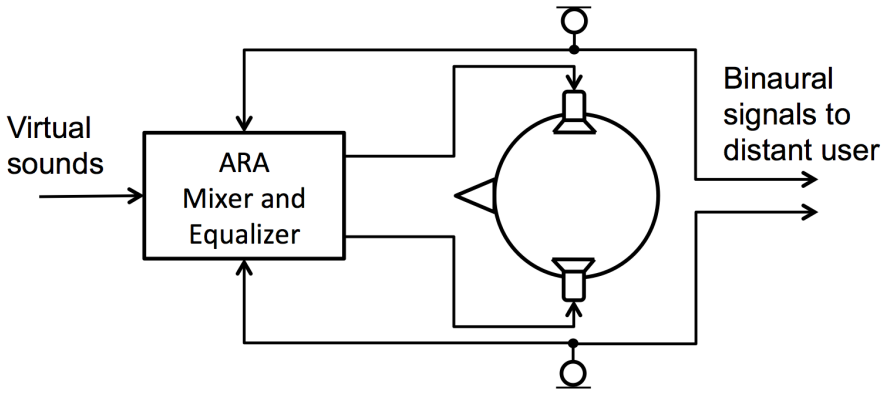

Figure 8: Block diagram of the ARA system. From [9].

2.5.2) [50]. In-ear headphones are a good alternative for the ARA headset, because they are already widely used and they often have good isolation characteristics [49]. Good isolation offers better control of leaked sound. Nevertheless, low frequencies may be problematic, because they leak more easily through and around the headset [50, 51, 52]. This leakage can be caused by tissue and bone-conduction, and sound may also propagate between the ear canal wall and the headphone cushion. Especially below 1 kHz the leaked sound may colour the perceived sound and thus deteriorate the pseudoacoustic representation [15].

Since headphones change the acoustics of the ear (see Section 2.3), equalization is needed to make the ARA system transparent [50]. This equalization is done in the ARA mixer, which can also be used to route the virtual sounds to the headphones or send the binaural signals to distant user for communications purposes. Low-latency requirements lead previously to analog realizations of the mixer and equalization [50]. However, analog components are bulky and expensive, and consequently there was interest for digital implementation of the equalization [15]. Digital realization offers programmability, ease of design and precision, but also introduces more latency. The effects of this latency can be minimized with the use of low-latency DSP-boards and adequate headset isolation. Rämö *et al.* [15] showed that digital realization is feasible and results in sufficiently good sound quality. They achieved a latency less than 1.4 ms, which is suitable for ARA system.

ARA systems can produce multiple problems that may deteriorate the user experience. One significant problem is connected to binaural cues and localization, which can be affected negatively [49]. Localization ensures situational awareness [53] and therefore it should be maintained when using ARA systems. If the binaural microphones are positioned at the ear canal entrance, spatial information is captured [53]. Hammershøi *et al.* [54] found that when the binaural microphones are either inside the ear canal, at the ear canal entrance or at most 6 mm outside the entrance the transmission from the microphone to the eardrum is independent of direction, i.e. the spatial information is already included in the recording

Additional problems may be caused by the headset cable [7]. Either it can

transmit mechanical noise to the ear canals or it can get stuck for example to clothing and thus cause nuisance. The solution to the noise is to use softer cable, but both problems may be solved in the future, when the ARA headset can be realized wirelessly. Furthermore, when ARA headset user is speaking, the occlusion effect occurs and own voice can be localized inside one's head [7]. Luckily, people seem to get accustomed to these phenomena quickly, so they do not cause much annoyance. Finally, wearing the ARA headset may cause social pressure, if other people do not realize that the ARA system user is able to properly hear them even with the headset on [51]. Despite all the problems, Tikander [51] found that ARA system is applicable to everyday life situations.

2.5.1 ARA Applications

The purpose of ARA systems is that they are used continuously long periods of time [7]. Therefore the best possible sound quality and transparency are important factors in the approval of ARA headset. However, users also need further motivation and added value to wear the headset, because pseudoacoustics representation alone is not motivational enough [51]. Therefore the usefulness of ARA systems comes with applications [51, 15, 55]. For that reason, the applications of ARA are discussed here shortly even though they are not the main object of this thesis.

The ARA applications can be divided in many different ways: for example there are communication and information applications or some applications can be seen as human-to-human and others human-to-machine communication [55]. Examples of communication applications that also can be categorized as human-to-human communication are full audio quality binaural telephony and teleconferencing [7, 55]. With ARA systems these application can have better telepresence than without it, since remote user and near-end user both hear the same pseudoacoustic environment.

Many applications also utilize position and orientation information of the user [55]. Global Positioning System (GPS) can for example provide the position of the user and head tracking can be used to acquire the orientation [56]. Such applications include eyes-free navigation [57, 58, 59], virtual tourist guides [51] and Auditory Sticker application [55]. The presented examples can also be classified as information applications or human-to-machine communication.

The functions of ARA system are often compared to assistive listening devices [7, 51, 53, 55, 60]. When allowing modification of the pseudoacoustic representation, it can be seen that ARA headset can be utilized as hearing aid device and intelligent earplug, or it can be used to emphasize important sound signals, such as alarms and warnings [49]. Recently, Rämö *et al.* [61] introduced a live sound equalization and attenuation system. It utilizes the ambient isolation of headphones to protect the hearing in concerts and the binaural microphones together with the ARA mixer to enable a real-time equalization of the concert. Thus, the system offers hearing protection, improves music quality and even improves speech intelligibility when used in noisy environments.

2.5.2 Comb-Filtering Effect

The comb-filtering effect occurs, when a signal is summed with a delayed version of itself [62, 63]. It produces audible comb-filter-shaped linear distortion, since a delay results in a linear phase shift across the frequency spectrum, which, in turn, leads to cancellation of some frequencies and reinforcement of others during the summation. The results of the comb-filtering effect depend on the amount of delay in the other signal as well as the amplitude difference of the two summed signals [63], though attenuation has greater effect of the two [62].

In an ARA system, the pseudoacoustic representation is the delayed signal, which is summed with the leaked sound at the ear drum. The delay is caused by the equalization filtering and especially by the analog-to-digital (AD) and digital-to-analog (DA) converters in digital ARA systems [9]. With low-latency DSP-boards the delay of these converters can be for example 0.95 ms, which is small enough for ARA systems. Furthermore, since the passive isolation of the headphones acts as a lowpass filter, it produces additional group delay for low frequencies [64]. This additional group delay can be 0.7–1.7 ms and thus allows a larger electronic delay without much deterioration in the sound quality.

The comb filtering effect in hear-through systems can be simulated with the following FIR comb filter [9]

$$H(z) = g_d + g_{eq} z^{-L_D}, \tag{6}$$

where g_d is the gain for the leaked sound, g_{eq} is the gain for pseudoacoustic representation and L_D is the delay in samples. Figure 9 shows the magnitude response of two different comb-filtering situations: in both examples the delay is 2 ms, but the gains of the signals differ. In Fig. 9(a) a worst-case scenario is shown, where both signals have the same amplitude, i.e. $g_d = g_{eq}$. As can be seen, the signals cancel each other out at the dip frequencies. However, in Fig. 9(b), the headphone isolation is taken into account. The isolation causes the leaked sound to attenuate, which decreases the comb-filter effect. When the attenuation is 20 dB, the depth of the dips is decreased to only 2 dB.

The audibility of the comb-filtering effect in relation to the delay and the amplitude difference has been studied e.g. in [65, 66, 62]. It has been found that also the type of sound signal affects the required attenuation. For example, Schuller *et al.* [66] reported that with small delays under 10 ms, an attenuation of 26-30 dB is needed to make the comb-filtering effect inaudible, when averaging over widely used audio test signals. On the other hand, Brunner *et al.* [62] studied the required attenuation for three different test signals: speech, piano and snare drum. They discovered that listeners were particularly sensitive to timbral changes in noisy signals, i.e. the snare drum signal. The average attenuation needed with the speech and piano signals were 12 dB and for the snare drum signal 17 dB, whereas the maximum values from a single subject were 22 dB for the piano, 23 dB for the speech and 27 dB for the snare drum.

The amount of isolation required for the comb-filtering effect to become inaudible cannot be reached with all kinds of headphones. However, as shown in Fig. 7 the

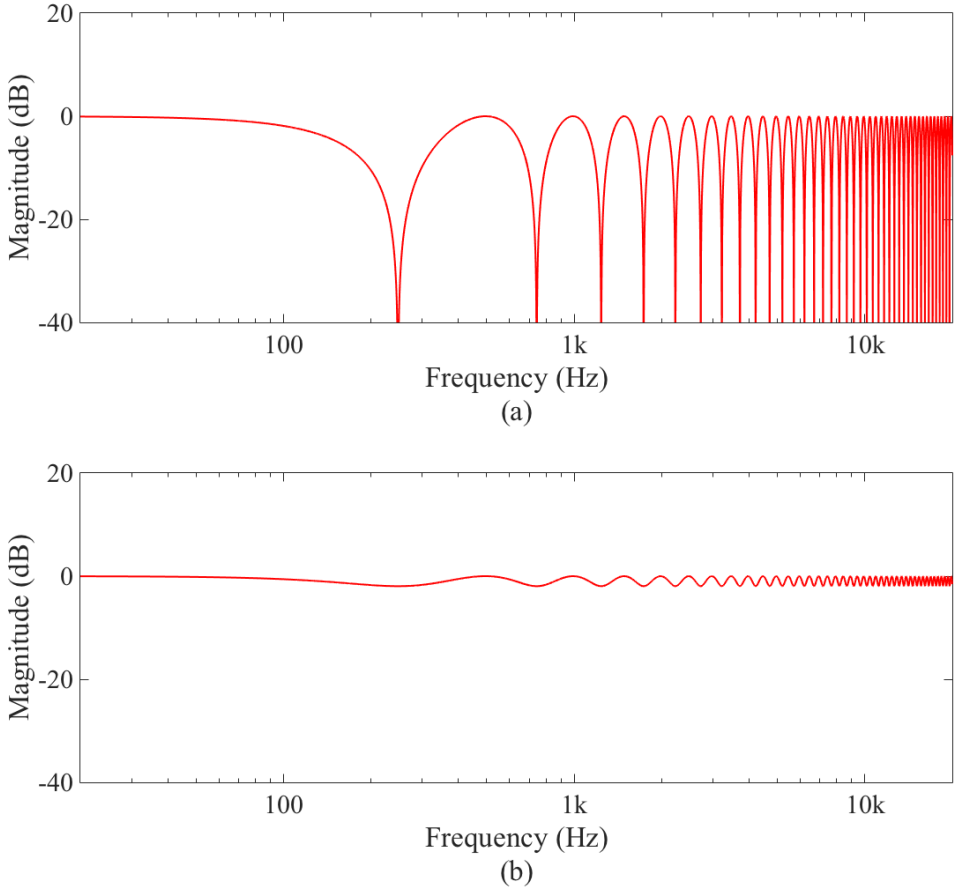

Figure 9: Magnitude frequency response of the FIR comb filter from the Equation (6). In both (a) and (b) the delay is 2 ms, in (a) $g_d = g_{eq} = 0.5$ and in (b) $g_d = 0.1$ and $g_{eq} = 0.9$.

headset used in this work has good isolation qualities, since it can provide more than 20 dB attenuation above approximately 900 Hz. Additionally, the fundamental principle of ARA system also diminishes the possible comb-filtering effect: since the ARA systems are designed to be worn long periods of time, the users are not able to compare the pseudoacoustic representation with the real world constantly, and thus they may adapt to slight timbral coloration [15].

3 Digital Filters and Filter Design

In this section digital filtering and many of its applications are discussed. They form the basic tools and techniques used in this thesis work, and thus this theoretical discussion gives background information for the next sections.

The structure of the section is as follows: First, the basic theory of digital filters is presented together with a specific application. Then, simple digital equalizers are discussed that are based on the parametric equalizer design. After that adaptive filtering is outlined, and especially the most widely used adaptive algorithm, namely least mean square (LMS) algorithm and its extensions, are discussed. The next subsection presents linear prediction, which is extensively used in speech processing, but nevertheless has applications in many other fields as well. The final two subsections recapitulate frequency warping and equivalent rectangular bandwidth (ERB) weighting, respectively. These both can be used to incorporate the effects of human auditory system and especially its nonuniform frequency resolution into DSP algorithms and thus improve their performance in human auditory sense.

3.1 Digital Filters

Filters are used to modify the magnitude and phase response of an input signal [67]. Analog filters use electronic components, such as resistors and capacitors, to operate on continuous-time signals, which also have continuous signal values [68]. Digital filters, on the other hand, operate on digital signals, which have both discrete time and discrete signal values [69]. When digital filters are used on analog sound signals, these signals have to first be sampled and quantized, i.e. transformed from voltage to numbers. This is achieved with an analog-to-digital (AD) converter. First, the signal is sampled, which converts it from continuous-time to discrete-time, and then quantization transforms continuous signal values to discrete ones. These steps have to be preformed exactly in this order, or otherwise gross errors will occur during the AD conversion [69]. After the digital filtering, the signals have to be converted back to continuous-time signals, and for that digital-to-analog (DA) converters are used.

Digital filters are often considered in two categories, which are finite impulse response (FIR) and infinite impulse response (IIR) filters. FIR filters only have feedforward structure, whereas IIR filters also have feedback structure [10]. This results in some differences. A FIR filter is straightforward to implement, since the filter coefficients directly form the impulse response, which is of finite length. The feedforward structure also ensures that FIR filters are always stable [67]. The feedback structure in IIR filters results, in theory, in an infinite impulse response. IIR filters may become unstable, and thus more care is needed when designing them [10]. However, IIR filters are not computationally as expensive as FIR filters, and they can also have smaller latency than FIR filters [67]. Finally, while FIR filters may be linear phase, IIR filters always produce some phase distortion to signals [10].

Digital filters have multiple advantages when compared to analog filters. Firstly, digital filters are programmable [15]. If one wants to alter the function of an analog filter, changes must be made to the hardware. With digital filters one can

just program new parameters to filters and thus change their behavior more easily. Secondly, digital filters can be created more precise and strict than analog filters [15], and therefore obtain better results from the filtering. Finally, adaptive filters can also be implemented straightforwardly with digital filters [70] (see Section 3.3).

However, digital filters have also disadvantages when compared to analog ones, and the biggest one is increased latency, i.e. the time between input and output. In analog filters electric signals propagate almost instantly through the filter and produce little latency. In digital filters, on the other hand, both the filtering itself and the AD/DA conversions cause latency. During AD conversions two types of latencies can occur: cycle latency and latency time [71]. The cycle latency is defined as the number of complete data cycles between the initiation of the signal conversion and the availability of the corresponding data. The latency time is from the beginning of the signal acquisition to the moment that the fully settled data is available to be read from the converter.

Digital filters usually have frequency dependent delays that cause latency. These can be illustrated with two alternative forms of the filter's phase response. The phase response, or the angle of the frequency response, $\Theta(\omega)$ gives in radians the phase shift that each sinusoidal input component will be subjected to [68]. From this, both phase delay and group delay can be derived. The phase delay gives the delay of each input sinusoid component in seconds, and it can be defined:

$$P(\omega) = -\frac{\Theta(\omega)}{\omega}. \tag{7}$$

The group delay on the other hand denotes the time delay of the amplitude envelope of a sinusoid at frequency ω, and it is defined as:

$$D(\omega) = -\frac{d}{d\omega}\Theta(\omega), \tag{8}$$

For linear phase systems these two delays are identical, but they differ from each other, if the phase response is nonlinear.

3.1.1 DC Blocker

One digital audio filter that is used in this work and will thus be presented here is a dc blocker. Often in signal processing applications, a constant-amplitude dc bias needs to be removed [72], and the dc blocker is a convenient algorithm for this. However, in this work the most attractive feature of the dc blocker is not related to the dc component, but instead it is used more like a simple and efficient high-pass filter.

The simplest way to block dc component would be to use the digital differentiator [72]. While it has infinite attenuation at 0 Hz, it also affects the components close to dc quite much, which is not desirable. This insufficiency can be solved with a cascaded leaky integrator, which is a nonideal integrator that leaks some energy away [72]. Together they give the following transfer function [72, 73]:

$$H(z) = \frac{1+p}{2}\frac{1-z^{-1}}{1-pz^{-1}}, \tag{9}$$

where p is a real pole ($0 < p < 1$). It represents a tradeoff between the bandwidth of the filter and its time-domain transient response. The first term is an amplification, which ensures that the dc blocker does not introduce positive gain to the signal.

As can be seen from Fig. 10, bigger p gives narrower and narrower notch at dc. While this is often a desired outcome, increasing the p also increases the duration of the impulse response [73]. Another way to affect the low frequency attenuation and the width of the notch at dc is to cascade two dc blockers together. This technique is advantageous in this work, since it offers increased attenuation in a narrow band near dc.

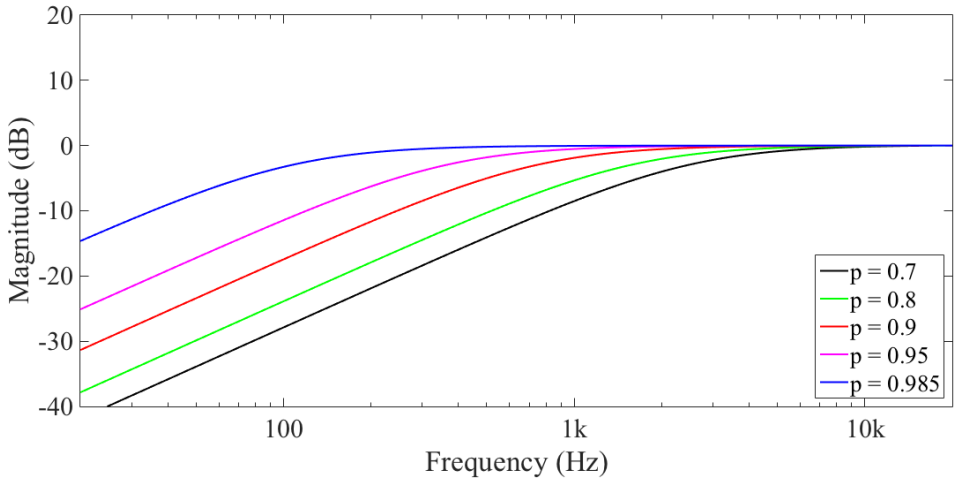

Figure 10: The response of dc blocker with different p values.

One possible drawback of the dc blocker is its nonlinear phase response, i.e. it causes different amount of group delay for different frequencies [72]. To counter this, the following trick can be used: If one has to correlating signals and uses dc blocker to another one, the correlation decreases. Luckily, an allpass filter can be used to the other signal to achieve similar group delay response, and thus restore the correlation. The allpass filter's transfer function is given by:

$$H(z) = \frac{z^{-1}-\lambda}{1-\lambda z^{-1}}, \tag{10}$$

where λ is the allpass filter parameter. When λ in Equation (10) is set to the same value as p in Equation (9), the group delay of the allpass filter is double that of the single dc blocker. But when two dc blockers are used as a cascade, together they cause equal group delay in relation to the single allpass filter. The group delay of a single and two cascaded similar dc blockers can be seen in Fig. 11(a), and the group delay of an allpass filter with the same parameter value can be seen in Fig. 11(b).

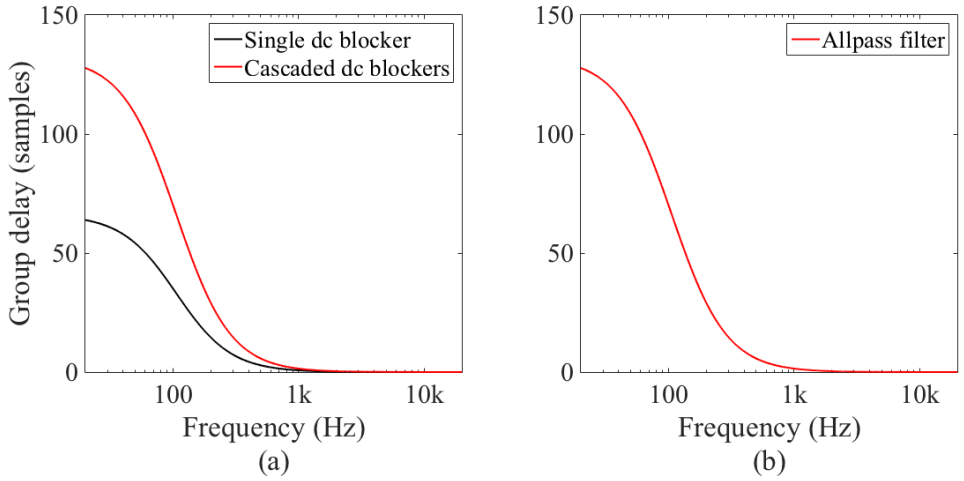

Figure 11: (a) The group delay of single dc blocker with $p = 0.985$ and two such dc blockers in cascade. (b) The group delay of single allpass filter with $\lambda = 0.985$.

3.1.2 Digital Resonator

Another digital audio filter that is used in this work is a resonator. A resonator has a peak in the amplitude response of the filter caused by a pole close to the unit circle [68], and thus it can be used as bandpass filter with a narrow bandwidth. A simple resonator has a two-pole structure, which, however, may lead to problems when the resonance frequency is set close to zero or the Nyqvist frequency [74]. Thus, an improved structure may be needed. It was introduced by Smith and Angell, who suggested that a feedforward section is added to the filter [74]. When the zeros are put at $z = \pm 1$, the resonator has the following transfer function [68, 74]:

$$H(z) = \frac{1 - z^{-2}}{1 - 2R\cos\theta z^{-1} + R^2 z^{-2}}, \tag{11}$$

where R is the pole radius and θ is the pole angle. When the center frequency ψ and bandwidth B of the filter have been chose in radians, R and the cosine of θ can be calculated as:

$$R = 1 - \frac{B}{2} \tag{12}$$

and

$$\cos\theta = \frac{2R}{1 + R^2}\cos\psi. \tag{13}$$

The frequency response of a resonator with center frequency of 1000 Hz and bandwidth of 100 Hz can be seen in Fig. 12.

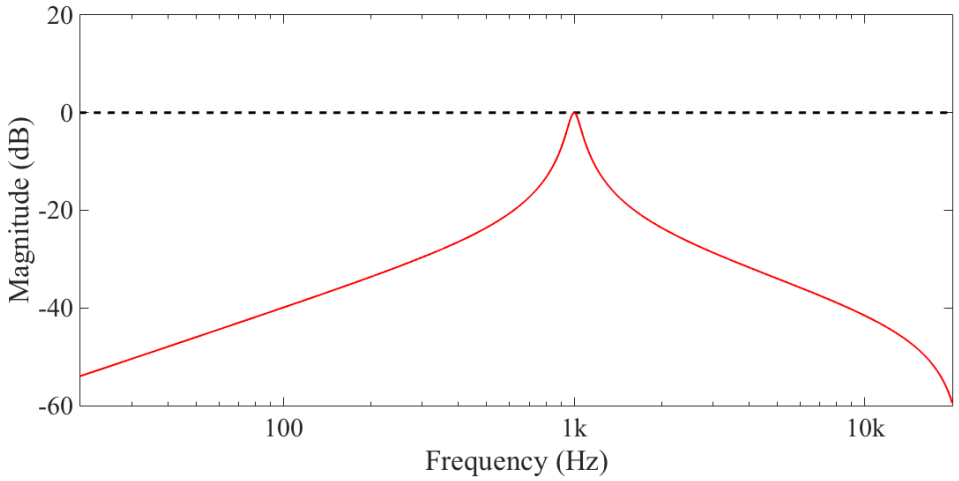

Figure 12: The frequency response of a resonator. The center frequency is 1000 Hz and bandwidth 100 Hz.

3.2 Digital Parametric Equalizers

Nowadays equalizers are often used for correcting or enhancing the performance of audio systems, i.e. the frequency response is altered to meet desired requirements [75]. For example headphones can be equalized to provide a natural music listening experience [18, 76, 77, 19, 20] or a natural hear-through experience in ARA applications [51, 15, 61, 78]. This differs from the original purpose, which was to flatten the frequency response [79]. For example, in telephone applications fixed equalizers were used to correct the audio transmission losses. In the 1930s variable equalizers were used for the first time for sound improvement, when poor sound reproduction systems were improved in motion picture theaters by equalizing them. In 1987 Yamaha introduced the first DSP-based, fully digital equalizer DEQ7.

Equalizers consist of set of filters [75]. Often these filters are either shelving or peaking filters [73]. Shelving filters are used to boost or cut low or high frequencies while leaving the remainder of the spectrum unaffected. The adjustable parameters for shelving filters are gain and cutoff frequency. Peaking filters, however, are used for midband equalization, and they can act as bandpass or bandstop filters. Gain can also be adjusted for peaking filters, and additionally they have center frequency and bandwidth parameters. Bandwidth can also be replaced with quality factor Q, which an inverse measure of the bandwidth [79] and characterizes the shape of the filter frequency response.

Two commonly used equalizer types are the parametric [80] and the graphic equalizer [81]. Parametric equalizers offer the control over all the parameters discussed in previous paragraph (i.e. the amplitude, the center frequency and the bandwidth), while in graphic equalizers user is only able to adjust the amplitude [79]. The name for graphic equalizers comes from the fact that they often have slide controls, and the slides then plot the approximate frequency response of the filter [73]. Next, tunable parametric equalizers are reviewed further, since they are utilized in this work.

The digital parametric equalizer introduced by Regalia and Mitra [80] is based on allpass filter structure. It allows individually adjustable equalizer parameters and also adjustable gain in certain frequency while leaving all the other frequencies unaffected. The Regalia-Mitra equalizer acts either as shelving or peaking filter depending whether first-order or second-order allpass sections are used respectively. Also, if $K > 1$, the equalizer is boosting shelving filter or a peak filter, and when $K > 1$, it becomes either cutting shelving filter or a notch filter.

The transfer function of the equalizer is:

$$H(z) = \frac{1}{2}[1 + A(z)] + \frac{K}{2}[1 - A(z)], \tag{14}$$

where K is the filter gain and $A(z)$ are the allpass sections [80]. For a shelving filter the allpass section is:

$$A(z) = \frac{z^{-1} + a}{1 + az^{-1}}, \tag{15}$$

where a is a frequency parameter:

$$a = \frac{\tan(\Omega_1/2) - 1}{\tan(\Omega_1/2) + 1} \tag{16}$$

and Ω_1 denotes the normalized filter cutoff frequency. The second-order allpass section, which is used for peaking filter, is:

$$A(z) = \frac{z^{-2} + b(1 + a)z^{-1} + a}{1 + b(1 + a)z^{-1} + az^{-2}}, \tag{17}$$

where

$$a = \frac{1 - \tan(\Omega_2/2)}{1 + \tan(\Omega_2/2)}, \tag{18}$$

$$b = -\cos(\omega_0), \tag{19}$$

Ω_2 defines the normalized filter bandwidth and ω_0 is the normalized center frequency.

However, the Regalia-Mitra equalizers have one problem, since they are asymmetric depending on whether positive or negative gain is used [80]. Zölzer and Boltze [82] proposed a small extension for the allpass structures, which offers a solution for this asymmetry. The solution concerns the frequency parameter a in both shelving and peaking filters: for first-order allpass section, the frequency parameter should become:

$$a = \begin{cases} \dfrac{\tan(\Omega_1/2) - 1}{\tan(\Omega_1/2) + 1}, & \text{when } \ K \geq 1 \\ \dfrac{\tan(\Omega_1/2) - K}{\tan(\Omega_1/2) + K}, & \text{when } \ K < 1 \end{cases} \tag{20}$$

and for the second-order situation it should likewise become:

$$a = \begin{cases} \dfrac{1 - \tan(\Omega_2/2)}{1 + \tan(\Omega_2/2)}, & \text{when} \quad K \geq 1 \\ \dfrac{K - \tan(\Omega_2/2)}{K + \tan(\Omega_2/2)}, & \text{when} \quad K < 1. \end{cases} \tag{21}$$

When these equations are used, the equalizers work symmetrically. The difference between asymmetrical and symmetrical behavior can be seen in Fig. 13.

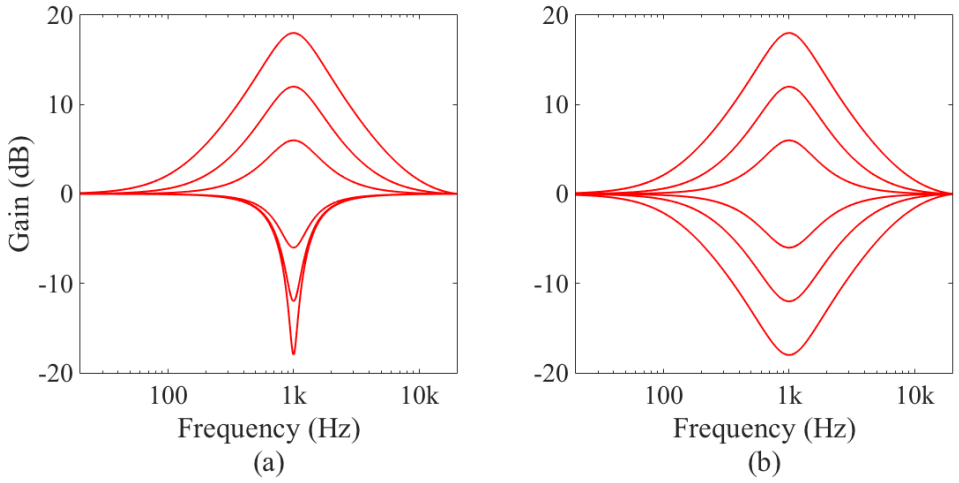

Figure 13: The effect of the asymmetry correction in Regalia-Mitra peaking equalizers, where in (a) the responses are of the original Regalia-Mitra design [80] and in (b) the correction by Zölzer and Boltze [82] is used.

3.3 Adaptive Filtering

Traditional digital filters presented in Section 3.1 are time-invariant, i.e. their coefficients are fixed, and in order to design them, one requires prescribed specifications [83]. However, in many real-world applications the processed signals may be time-varying or the required specifications may be unknown, and thus the fixed filter coefficients are not optimal [67]. Instead, adaptive filters can be used. An adaptive filter is a time-varying filter, which modifies its coefficients recursively in accordance with the input data [84]. Since adaptive filters use the input signal to adjust the filter coefficients, they are able to adapt to changing conditions and to learn the characteristics of the input signal [67].

The basic schematic diagram of an adaptive filter can be seen in Fig. 14. An input $x(k)$ is fed to the adaptive filter, which results in an output $y(k)$. This output is then compared to the desired signal $d(k)$ and error is calculated $e(k) = d(k) - y(k)$. Finally, the filter parameters are adjusted with the help of the error signal and a rule, i.e. adaptive algorithm [83]. The basic principle of the algorithm is to minimize the chosen objective function. One possible choice for the objective function is the mean-square error (MSE): $E[|e(k)|^2]$ [83]. In an FIR adaptive filter realization MSE

yields a quadratic function, which has an unique optimal solution, and thus a point to which the adaptive filter tries to converge.

When comparing the performances of different adaptive algorithms, several factors are important, such as [84]:

- Rate of convergence, which is defined as the number of algorithm iterations needed to converge to the vicinity of the optimum solution in the mean-square sense.
- Misadjustment, i.e. the steady-state error of the algorithm. It is obtained by comparing the output of the adaptive algorithm to the optimum solution.
- Tracking ability, which denotes, how well the algorithm tracks signal variations in a nonstationary environment.
- Robustness, i.e. sensitivity to small disturbances that can only result in small estimation errors,
- Computational requirements, which consist of the number of operations required for one iteration, the size of memory locations required to store the data, and the investments required to program the algorithm.
- Numerical properties, i.e. numerical stability and numerical precision in the presence of quantization errors.

These performance factors are often conflicting and as a consequence, compromises are needed: for example, a fast convergence rate can result in either larger misadjustment or more demanding computational requirements.

As stated previously, adaptive filters are more suitable to certain applications than traditional fixed filters. The adaptive filter applications are often divided into four basic classes depending on the manner which is used to extract the desired response [84]. These four groups are system identification, inverse modeling, signal enhancement, and prediction [83, 84]. In system identification, the adaptive filter is used to obtain a model of an unknown system. Same input signal is fed to both the unknown system and the adaptive filter, and the output of the former is used as a desired signal for the latter. Thus, when the filter is converged, it represents the input-output response relationship of the unknown system. The system identification is the most interesting application group, since in this thesis work adaptive filter is used for modeling a dynamic system.

The inverse modeling is quite similar to the system identification, but here the purpose is to obtain an inverse model that can be used to equalize a system [83, 84]. For example, when improving dispersive communications channel, the transmitted and corrupted signal is used as the input and a delayed version of the original signal is used as the desired signal. The signal enhancement can be used, for example, to cancel the power line interference in electrocardiography [83, 84]. If a signal is corrupted by noise, which can be measured alone, the noise signal can be used as the input and the corrupted signal as the desired signal. After the convergence, the

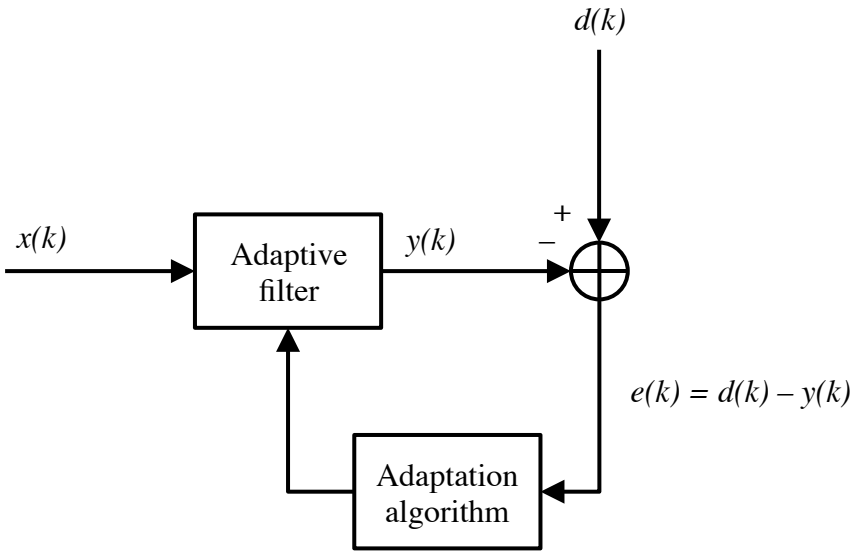

Figure 14: Schematic diagram of an adaptive filter.

output error will be an enhanced version of the original signal. Finally, in prediction a delayed signal is fed to the adaptive filter while using the original one as the desired signal [83, 84]. This forms a predictor model for the signal, and it can be utilized for example in adaptive line enhancement.

3.3.1 Least Mean Square Algorithm

From the various adaptive filter algorithms, LMS (Least Mean Square) algorithm is the most popular one [83, 84]. It was introduced in 1960 by Widrow and Hoff [85], and since then it has been widely used due to its computational simplicity and effectiveness [73, 83, 84, 86]. LMS is a gradient-based algorithm, i.e. it uses instantaneous values as an estimate for the gradient vector of the objective function to determine the direction of adjustment for the filter coefficient that leads to optimum solution in mean-square sense [83, 84]. Due to using estimate of the gradient vector, LMS is always noisy [83, 86].

According to the LMS algorithm, the equation for coefficient update is:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu e(k)\mathbf{x}(k), \tag{22}$$

where $\mathbf{w}(k)$ is the filter coefficient vector, μ is the step size, $e(k)$ is the error signal, and $\mathbf{x}(k)$ is the input vector for the filter ($e(k)$ and $\mathbf{x}(k)$ can be seen in Fig. 14). The error is calculated by:

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k), \tag{23}$$

where the latter term is the filter output from the adaptive filter. Often the filter

coefficient vector is initialized to zero [83], but some estimate values can also be used as initial values. The algorithm then uses Equation (22) to minimize the error $e(k)$ by alternating between filtering process and adaptation process [84].

The step size μ is essential when it comes to the performance of the LMS algorithm due to its effects on stability, convergence speed, and misadjustment [84]. When quick convergence rate is desired, step size can be chosen to be large. However, this results in large misadjustment, and if step size is chosen too large, the algorithm will not converge. The stability condition for the step size is [84]:

$$0 < \mu < \frac{1}{\mathrm{tr}[\mathbf{R}]}, \tag{24}$$

where $\mathrm{tr}[\mathbf{R}]$ is the trace of autocorrelation matrix $E[\mathbf{x}^T(k)\mathbf{x}(k)]$. On the other hand small step size leads to small misadjustment, but also to slow convergence rate. This problem can be solved with the NLMS (Normalized Least Mean Square) algorithm.

3.3.2 Normalized Least Mean Square Algorithm

NLMS algorithm is an extension to LMS algorithm that was suggested independently by both Nagumo and Noda [87] and Albert and Gardner [84] in 1967. NLMS has the same structure as LMS and only introduces a time-varying step size [84]. The new step size is normalized by an estimate of the input signal power:

$$\mu_{NLMS} = \frac{\mu}{\mathbf{X}^T(k)\mathbf{X}(k) + \epsilon}, \tag{25}$$

where $0 < \mu \leq 2$ and ϵ is a small constant that prevents the division with zero when input signal is really small. With this change the convergence rate of the algorithm improves, because the variable step size results in variable convergence factor and minimization of the instantaneous output error [83]. However, both LMS and NLMS have similar drawback since they converge slower with colored noise than white noise input [86].

3.3.3 Robust Variable Step Size Normalized Least Mean Square Algorithm

Additional extensions to the LMS and NLMS algorithms were introduced by Vega *et al.* [88], who named their algorithm the RVSS-NLMS (Robust Variable Step Size Normalized Least Mean Square) algorithm. It is based on the optimization of the square of the *a posteriori error*, which is defined as:

$$e_p(k) = \mathbf{x}^T(k)\mathbf{w}(k) + \upsilon(k), \tag{26}$$

where $\upsilon(k)$ is the noise that corrupts the filter output $y(k)$ consisting of both background measurement noise and impulsive noise.

In real-world applications there are often perturbations that deteriorate the performance of adaptive filters [88]. If a large noise sample is present in the input signal, large change occurs in the adaptive filter coefficients, which leads to increased

error and degradation of performance. Thus, a robust algorithm is beneficial. Here, robust is used as "slightly sensitive to large perturbations" [88]. This is ensured by restricting the energy of the filter update at each iteration as follows:

$$\|\mathbf{w}(k) - \mathbf{w}(k-1)\| \leq \delta(k-1), \tag{27}$$

where δ is a positive sequence [88]. In practice the robustness of the RVSS-NLMS algorithm is based on the automatic choice between two operating modes.

The RVSS-NLMS algorithm can be formulated as [88]:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \min\left[\frac{|e(k)|}{\|\mathbf{x}(k)\| + \epsilon}, \sqrt{\delta(k-1)}\right] \text{sign}[e(k)] \frac{\mathbf{x}(k)}{\|\mathbf{x}(k)\| + \epsilon}, \tag{28}$$

where ϵ is small constant and the error is calculated:

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k-1). \tag{29}$$

Depending on the min-function, this algorithm acts either as a NLMS (see Section 3.3.2) or a NSA (Normalized Sign Algorithm). Sign algorithm is quite similar to LMS algorithm, but instead of the error signal in Equations (22) only the sign of that error is used [83, 89]. However, another way to interpret the algorithm is as another NLMS with variable step size [88]

$$\mu = \min\left[\frac{\sqrt{\delta(k-1)}}{|e(k)|/\|\mathbf{x}(k)\|}, 1\right]. \tag{30}$$

As a result of these two operating modes, the algorithm can have both the fast convergence rate of NLMS and the robustness of NSA against noise [88].

In the selection of the delta sequence same principles are used as in the selection of the step size in LMS algorithm: initially the values should be as large as possible for high convergence rate and during the steady-state performance it should have lower values to ensure lower error [88]. Therefore, δ is defined as:

$$\delta(k) = \alpha\delta(k-1) + (1-\alpha)\min\left[\frac{e(k)^2}{\|\mathbf{x}(k)\|^2}, \delta(k-1)\right], \tag{31}$$

where α is a memory factor and $\delta(0)$ can be selected as:

$$\delta(0) = \frac{\sigma_d^2}{\sigma_x^2 M}, \tag{32}$$

where M is the filter order, σ_d^2 and σ_x^2 are the desired signal and input signal power, respectively. The memory factor α can be chosen according to the rule:

$$\alpha = 1 - \frac{1}{\kappa M}, \tag{33}$$

where parameter κ depends on the color of the input signal (typically $1 \leq \kappa \leq 6$). Both the memory factor α and $\delta(0)$ control the tradeoff between convergence rate and robustness [88]. When the delta sequence is defined like this, it converges towards zero. Additionally, Vega *et al.* proved that the limiting mean-square misadjustment is zero under some reasonable assumptions [88].

3.4 Linear Prediction

Linear prediction (LP) is a method to predict the current signal sample from a linear combination of the previous samples of the signal so that the predicted value is the best one in the least mean square sense [90]. The mathematical representation for this is:

$$\hat{s}(n) = \sum_{k=1}^{N} a_k s(n-k), \tag{34}$$

where a_k is the predictor coefficient, N is the order of the analysis, $s(n)$ is the original signal and $\hat{s}(n)$ is the predicted value. Linear prediction can be used to represent the spectrum of a signal [94]. With LP, the spectral information can be compressed into few filter coefficients, i.e. the coefficients a_k. The first known application of LP is the analysis of sunspots [91] and it is still widely applied in many signal processing fields [90]. It is especially powerful technique in speech and audio signal processing: for example, LP is a basic method for speech compression and coding [92] and also for speech and speaker recognition [93].

There are multiple ways to calculate the predictor coefficients a_k, but the most frequently used one is the autocorrelation method [67]. In autocorrelation method the coefficients are calculated with a linear matrix equation, which can be formulated as:

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_p \end{bmatrix}, \tag{35}$$

where r_k are the autocorrelation coefficients:

$$r_k = \sum_{i=-\infty}^{\infty} s(i)s(i+k). \tag{36}$$

Since the matrix in Equation (35) is a symmetric Toeplitz matrix (i.e. all the elements along each diagonal are equal), the solution can be obtained quickly, for example with the Levinson-Durbin recursion [95]. When the equation is solved for the predictor coefficients, the all-pole synthesis filter $1/A_{inv}(z)$ is obtained, where [67]:

$$A_{inv}(z) = 1 - \sum a_p z^{-p}. \tag{37}$$

The all-zero filter $A_{inv}(z)$ is called the inverse filter [90]. When this inverse filter is used on the original signal, a spectrally flattened, i.e. whitened, residual signal is obtained [67]. This property of the LP is used in this work.

3.5 Frequency Warping

Frequency warping means designing or implementing DSP algorithms directly on a warped, i.e. nonuniform, frequency scale that is relevant for auditory perception of the human auditory system [96]. The nonuniform frequency scale is achieved by replacing the unit delays of digital filters with first-order allpass filters. Based on filter design method by Constantinides [97], Schüßler proposed this replacement to produce frequency-warped transfer functions [96]. Oppenheim *et al.* [98, 99, 100], on the other hand, used frequency-warping together with the FFT to produce nonuniform spectral representation of signals, and Strube [101] showed that the warping effect can be adjusted to approximate the spectral representation of the human auditory system. Finally, Härmä *et al.* [96] wrote a comprehensive paper on the frequency warping and its audio applications.

The conventional techniques to design DSP algorithms result in linear frequency scale in relation to the hertz scale [96]. This is due to the unit delay z^{-1}, which delays all the frequency components equal amount. First-order allpass filters (see Equation (10)), on the other hand, produce frequency-dependent delay [96] while passing all frequencies through with equal gain [73], and therefore they can be used to achieve nonuniform frequency scales.

The phase response, and thus the group delay, of the allpass filter can be adjusted with the parameter λ [96]. When using positive values for the parameter λ in allpass filter chains, low frequency components have much higher group delay than high frequency ones, and therefore they propagate slower through the filter chain. Thus, the allpass filters in frequency warping form a dispersive system and produce frequency-dependent resampling of the signal [96]. That is, frequency warping increases the frequency of each sinusoidal component and also changes the temporal structure of the original signal. This effect can be seen in Fig. 15, where both original and warped versions ($\lambda = 0.723$ and length of the allpass filter chain is 1000) of three sinusoidal signals and their sum spectrum can be seen.

In frequency warping, replacing the unit delays with first-order allpass filters can be interpreted as a bilinear transformation determined by the mapping [96]:

$$z^{-1} \rightarrow \tilde{z}^{-1} = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}. \tag{38}$$

Therefore the mappings between a traditional signal or impulse response $s(n)$ and the corresponding warped signal or impulse response $w(k)$ can be expressed as follows: a method to calculate a warped signal from the original one is:

$$\sum_{n=0}^{\infty} s(n) \left(\frac{\tilde{z}^{-1} + \lambda}{1 + \lambda \tilde{z}^{-1}} \right)^n = \sum_{k=0}^{\infty} w(k) \tilde{z}^{-k} \tag{39}$$

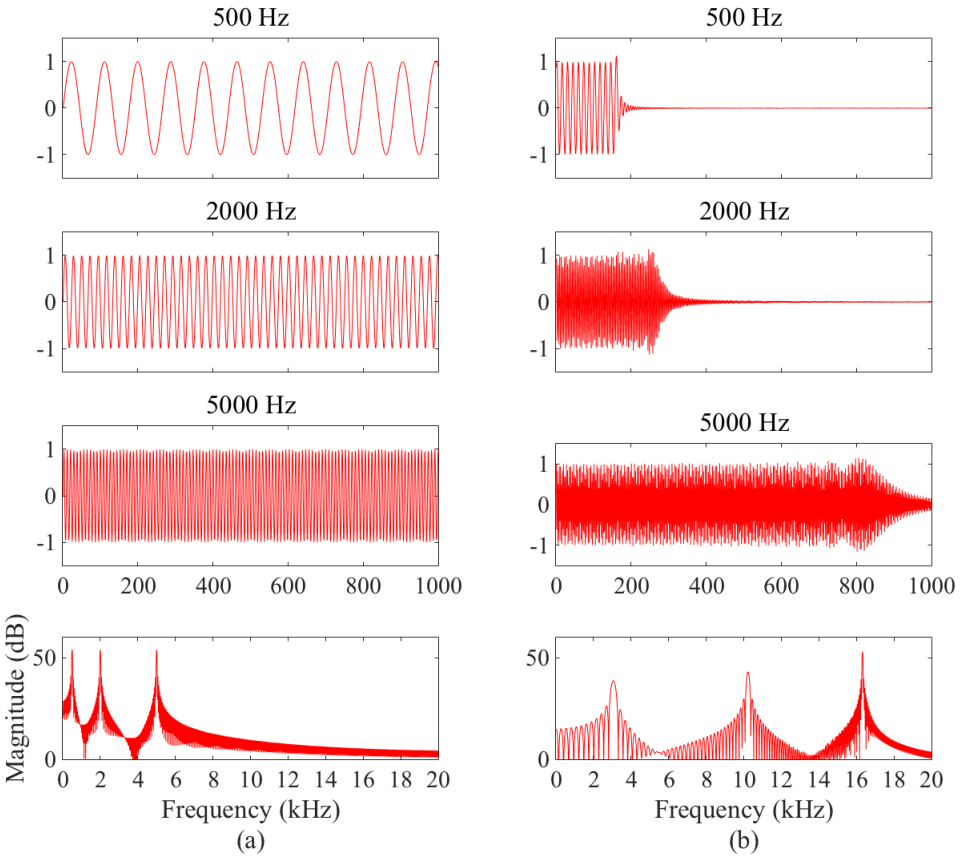

Figure 15: The effect of frequency warping: (a) A set of three sinusoidal signals in time-domain and the spectrum of their sum. (b) The same signals frequency warped and the warped spectrum.

and another method to synthesize the original signal from its warped counterpart is [96]

$$\sum_{n=0}^{\infty} s(n)z^{-n} = \sum_{k=0}^{\infty} w(k)\left(\frac{z^{-1}-\lambda}{1-\lambda z^{-1}}\right)^{k}. \tag{40}$$

As stated, the warping technique can be used either on a signal segment or a transfer function, but preferably the latter is used, since it is more straightforward [96].

Since frequency warping can be interpreted as a bilinear mapping from a unit disc onto another unit disc, it may be combined with most of the conventional DSP methods [96]. For example warped FIR and IIR filters can be directly designed with the conventional design methods, and most of the techniques for parametric spectral estimation, adaptive filtering, and predictive coding can be utilized immediately with warped systems, which makes it a highly useful tool. This fact together with the automatic utilization of nonuniform frequency scales are the main advantages of frequency warping. Furthermore, gains can be achieved in terms of quality and computational complexity [96], because frequency warping enables for lower order

filters that approximate the desired target well in auditory sense [102].

3.6 ERB Weighting

As was discussed in the previous section (see Section 3.5), the linear frequency scale used often in DSP algorithms is not optimal from the human auditory system point of view. In order to incorporate the effects of human auditory system and the nonuniform resolution into DSP algorithms, auditory weighting can be used [102]. One such technique is to use ERB (Equivalent Rectangular Bandwidth) weighting function to filter responses, since it has been found experimentally that ERB corresponds really well the auditory resolution [104, 67].

According to modern psychoacoustic theory, the human hearing can be modelled as a set of filters, which are called the auditory filters [103]. These auditory filters can be approximated as rectangular bandpass filters, whose bandwidth is characterized by ERB [67], i.e. the bandwidth of a rectangular filter that passes the same amount of signal energy as the corresponding auditory filter [96]. The ERB is calculated as [104, 103]:

$$\Delta f_{ERB} = 24.7 + 0.108 f_c, \tag{41}$$

where f_c is the center frequency. When the inverse of these calculated bandwidths is taken, the weight of the each frequency point is obtained and thus also the ERB weighting function [102]. The response of this function can be seen in Fig. 16, and when it is used to filter signals, the results correspond to the human auditory resolution.

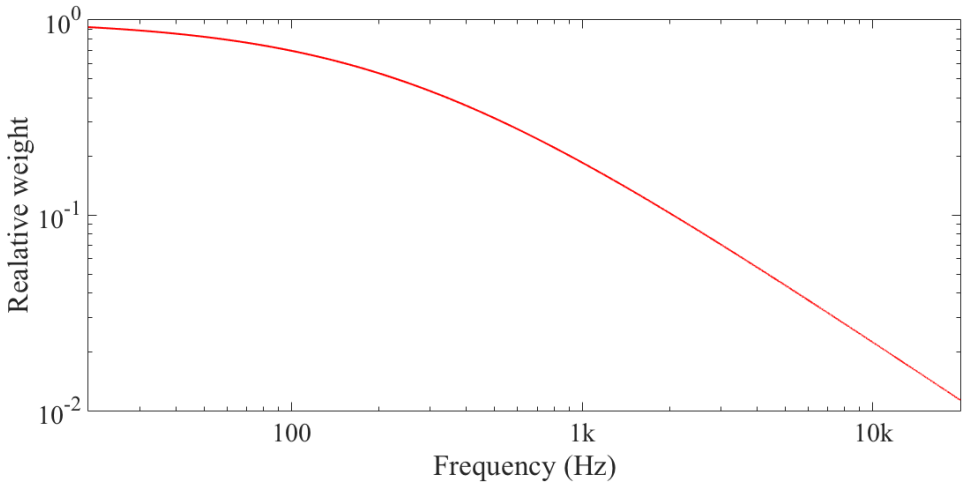

Figure 16: Auditory ERB weighting function as a function of frequency.

4 Adaptive Hear-Through Algorithm

In this section the final product of this thesis work will be presented. The product is a signal processing algorithm developed for Nokia to be used with a prototype headset. The proposed algorithm is implemented using Matlab and Playrec.

First, the main features of the prototype headset are described, since they regulate the starting point for the algorithm. Next, two distinct parts of the proposed algorithm are discussed, i.e. adaptive isolation estimation and hear-through isolation. Each part is divided into functional blocks and the behavior of each block is presented separately. The theory of these blocks is already discussed in the previous section, and therefore in this section only the practical behavior is recapitulated. Finally, the operation of the complete algorithm is described as well as the operation of the Matlab implementation.

4.1 Prototype Headset

The headset used in this thesis work was constructed specifically for this project and thus it is a prototype. The earpieces of the headset are insert-type and therefore offer good isolation. The transducers in the earpieces are electromagnetic, i.e. balanced armatures.

The headset can be regarded as an ARA headset since it has microphones outside both earpieces that can be used for ARA and hear-through applications. In addition the earpieces have second microphones close to the transducer. These microphones are thus inside the ear canal of the user when the headset is worn. By utilizing both microphones in one earpiece the isolation of that earpiece can be estimated: since one microphone is outside the ear canal and the other one is inside it their deconvolution is a transfer function that describes the attenuation of sound through the earpiece.

4.2 Adaptive Isolation Estimation

Figure 17 shows the block diagram of the proposed algorithm. Additionally, the two distinct parts, i.e. adaptive isolation estimation and hear-through equalization, can be seen as well as the blocks that form them. The functions are shown in the block diagram as mono systems, i.e. only the processes of one ear are shown. However, since the headset consists of two earpieces with two microphones in each earpiece, the proposed algorithm performs the same processing for both ears separately in the adaptive isolation estimation part. Thus a separate isolation estimate is obtained for both ears. In hear-through equalization, the ambient sounds are recorded binaurally so that the signal intended for right ear playback is recorded with the out microphone of the right earpiece and the signal for the left ear is recorded with the left out microphone. The hear-through equalization is identical for both recorded signals so that unnatural artifacts are avoided that could result in a different timbre in each ear. The only exception occurs, when only one of the earpieces is poorly inserted, and thus needs adjusted equalization.

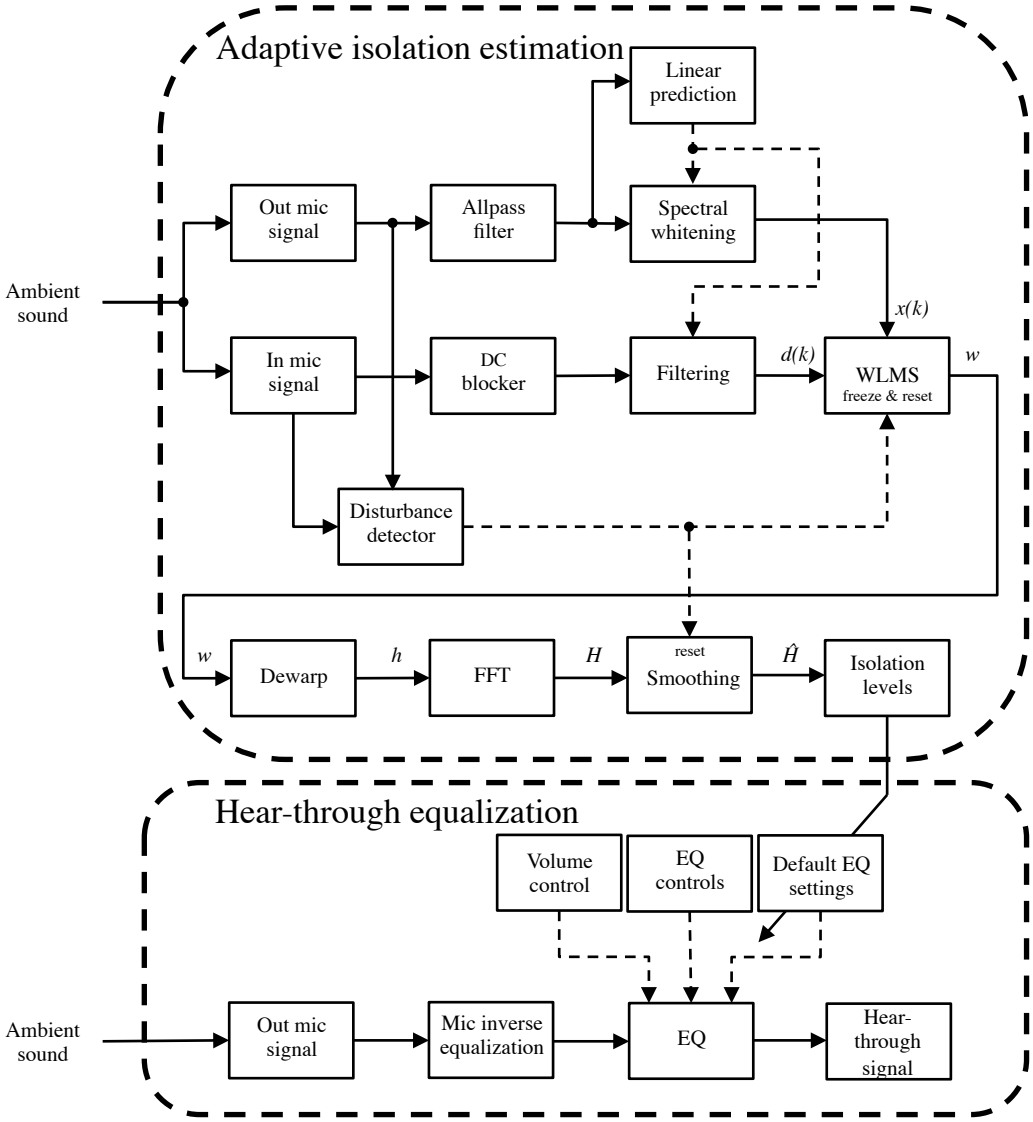

Figure 17: Block diagram of the proposed algorithm. The two separate parts are marked with dashed line in bold. Solid lines represent signals in time or frequency domain and dashed lines represent control signals for certain blocks.

In the following sections first the blocks of the adaptive isolation estimation are discussed and after that the blocks of the hear-through equalization are described.

4.2.1 Disturbance Detector

The first step of the proposed algorithm is to obtain the signal from both microphones: the outside microphone and inside microphone, as seen in Fig. 17. These signals will be designated here as out signal and in signal, respectively. Before these signals are processed in any other manner, they are passed to the disturbance detector, which compares the powers of the two signals.

In order to simplify the disturbance detector and reduce its computational cost, the whole spectra of both out and in signals are not processed. Instead a second order IIR resonator filter is used as a bandpass filter. The resonance is selected at 1000 Hz since neither much lower frequencies nor much higher frequencies are stable enough to be analyzed. The bandwidth of the resonator is 100 Hz. Since the highest signal frequency is reduced from 22050 Hz to approximately 1000 Hz with the resonator filter, the sample frequency can also be reduced. The reduction of the sample frequency is achieved by keeping only every tenth data point. After the lengths of both the out and in signals are reduced, their signal energies are calculated. The energy of a signal $x(n)$ can be calculated as $\hat{E} = \frac{1}{M}\sum_{k=0}^{M-1} x(n-k)$, where M is the length of the signal.

The disturbance detector uses the out and in signal energies in two ways. Firstly, the in signal energy is compared to the out signal energy, and if the former is larger, the proposed algorithm freezes until the in signal energy decreases. Since the headset attenuates the sounds propagating to the ear, the in signal energy should be smaller than the out signal energy.

However, if the user is for example talking while wearing the headset, the occlusion effect results in highly increased sound pressure inside the ear canal. This, in turn, leads to higher in signal energy than out signal energy and disturbances in the microphone recordings. Another cause for increased in signal energy is the wire noise that is conducted to the ear, which is caused by the casing of the wire rubbing against clothing. The freezing of the algorithm prevents vast processing errors that could follow from corrupted recordings.

Secondly, the proposed algorithm requires certain minimum threshold surpassed before it can function properly. Therefore, a minimum value for the in signal energy is defined. If the actual energy is smaller than the defined minimum value, the algorithm freezes until enough sound energy is present again. The in signal energy can decrease dramatically, for example, when moving from a street inside a building in a high traffic area.

4.2.2 DC Blocker and Allpass Filter

The first blocks that process the actual target signals are the dc blocker for the in signal and corresponding allpass filter for the out signal. The in signal is filtered with a dc blocker because the frequency band of the proposed algorithm is limited to 100–10000 Hz. Thus, the dc blocker acts as a high-pass filter that removes the unwanted portions of the signal below 100 Hz. The low limit of the frequency band is chosen since the in signal is partly corrupted by the heartbeat of the user. In addition, the upper limit is chosen since attempts to equalize audio signals above 10 kHz result in audible and distracting artifacts [50].

As the prototype headset used in this thesis work has insert-type earpieces, it is tightly connected to the ear canal. Thus, the frame of the earpiece can capture the heartbeat from the ear canal walls. The vibration of the earpiece frame is then, in turn, conducted to the in microphone of the earpiece and to the recordings. Such partly corrupted measurement can be seen in Fig. 18(a).

To improve the results of the proposed algorithm, the heartbeat should be removed from the in signal before applying any further processing. This is achieved by selecting the dc blocker parameter $p = 0.985$ and by cascading two identical dc blocker sections in succession. However, as discussed in Section 3.1.1, the dc blocker adds nonlinear group delay to the filtered signal. Therefore, also the out signal needs to be processed and a single allpass filter with the same filter parameter value $\lambda = 0.985$ is chosen to maintain the correlation of the in and out signals. The signal from Fig. 18(a) but this time filtered with the dc blocker can be seen in Fig 18(b).

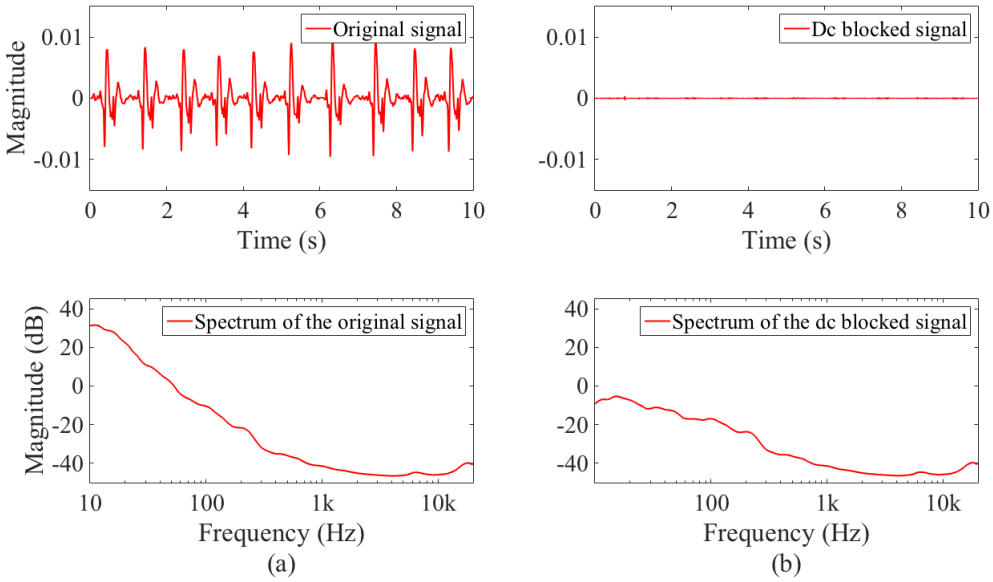

Figure 18: The heartbeat and its removal from the in signal. (a) An in signal recorded in silence and corrupted by the heartbeat of the user, and the spectrum of that signal. (b) The same signal filtered by the two cascaded dc blockers with the parameter $p = 0.985$, and the spectrum of that signal.

4.2.3 Linear Prediction and Spectral Whitening

After the dc blocker and the corresponding allpass filter, the next blocks in Fig. 17 are the spectral whitening for the out signal and the corresponding filtering for the in signal. These processes are performed to increase the convergence speed of the WLMS algorithm, since NLMS algorithms converge the fastest with a white noise input [86] and the used WLMS algorithm is a type of NLMS.

First, the out signal goes through a first order linear predictor that extracts the spectrum of that signal. Since a first order LP is used, the equations shown in Section 3.4 are simplified significantly. The one predictor coefficient a_1 is obtained from Equation (35) that is reduced to:

$$a_1 = \frac{r_1}{r_0}, \tag{42}$$

where r_1 and r_0 can be obtained from Equation (36) as an autocorrelation of the signal with the lag of one and zero samples, respectively. Now, an all-zero inverse filter $A_{inv}(z)$ can be written $A_{inv}(z) = 1 - a_1 z^{-1}$. When this filter is used to the out signal, its spectrum will be approximately whitened.

Since the inverse filter filter $A_{inv}(z)$ is obtained from the out signal, it can be used to whiten only the spectrum of that particular signal. However, likewise to the previous blocks, the correlation between the in and out signals needs to be maintained and thus also the in signal is processed with the same filter $A_{inv}(z)$. This may not result in a completely white signal spectrum for the in signal since the spectra of the in and out signals usually differ.

4.2.4 Warped LMS

The next block in Fig. 17 is the warped LMS which contains the most significant operations of the proposed algorithm. All the previous blocks are used to preprocess the recorded signals, and in the LMS block the warped RVSS-NLMS (WRVSS-NLMS) algorithm is used to estimate the momentary isolation from the preprocessed in and out signals. This block takes as the input the out signal, which is the signal $x(k)$ in Equation (29), and the in signal, which becomes the reference signal $d(k)$. The result of this block is an L-length vector containing filter coefficients that simulate the isolation properties of the headset.

First of all, the warping part is discussed. As stated in Section 3.5, frequency warping can be performed either on a signal segment or a transfer function. In the proposed algorithm the transfer function of the LMS algorithm is warped. However, in the Matlab realization of the proposed algorithm it can be argued that the out signal is warped before it is fed to the LMS algorithm.

The frequency warping is used in the proposed algorithm, because it enables reduced LMS algorithm length. Since the human auditory resolution emphasizes low frequencies, it is necessary to model the low frequency behavior of systems accurately. Normally this would require higher filter orders, but when frequency warping is used, these low frequencies are automatically emphasized and filter order can be maintained low. In the proposed algorithm, this reduction of the filter order also outweighs the increase in computational cost caused by complexity of allpass filters when compared to cascaded unit delays.

In frequency warping, a turning point frequency determines the point frequency, which is not affected by the warping [96]. However, the frequencies below that are emphasized, while the frequencies above are de-emphasized. The turning point can be calculated as follows [96]:

$$f_{tp} = \pm \frac{f_s}{2\pi} \cos^{-1}(\lambda), \tag{43}$$

where f_s is the sampling frequency and λ is the warping parameter (i.e. the allpass filter parameter). In the proposed algorithm, a value of 0.7 is chosen for the parameter, which results in $f_{tp} \approx 5600$ Hz. This means that the frequencies below 5600 Hz are emphasized while frequencies above it are de-emphasized.

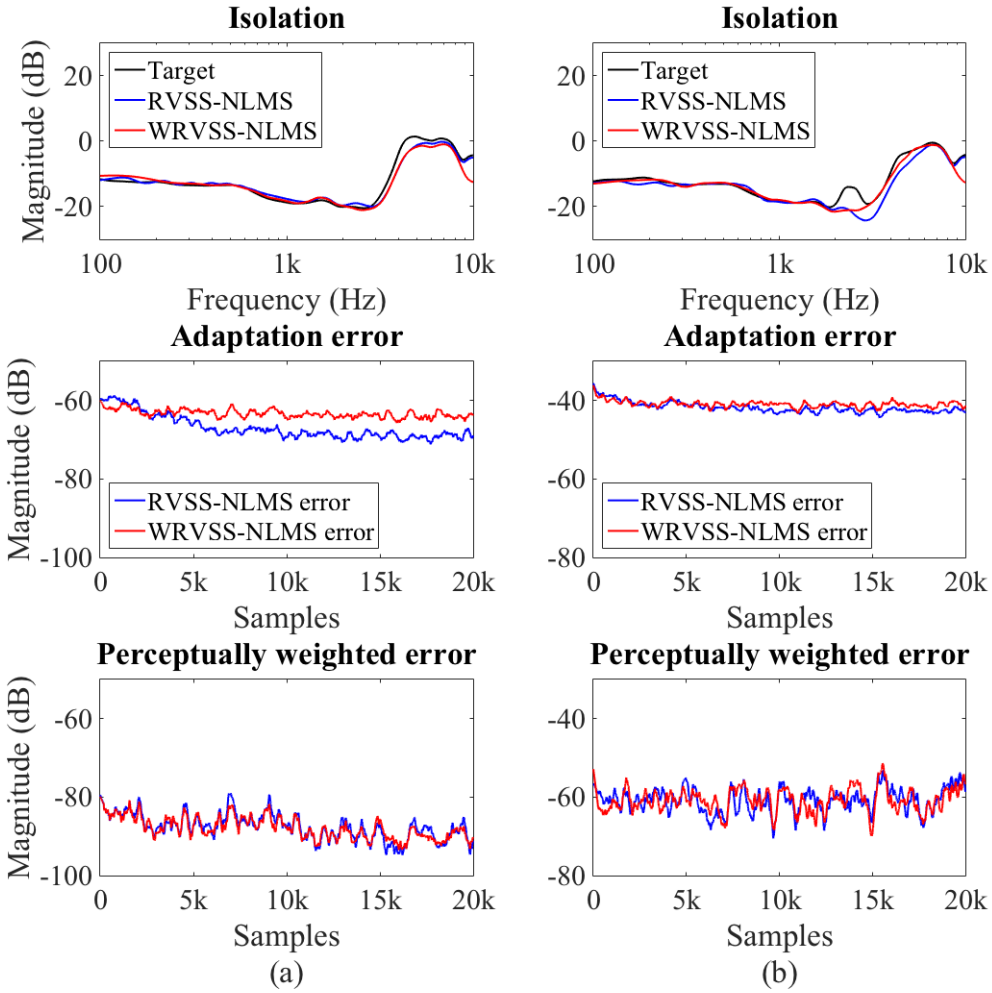

Figure 19: The effect of frequency warping on RVSS-NLMS algorithm. The topmost figures show the isolation estimation in comparison to target isolation, the middle figures show the beginning of the adaptation error signal, and the bottom figures show perceptually weighted error signals in their entirety. In (a) the algorithm is tested with white noise and in (b) with pinkish simulated bus noise as an external noise source. For every RVSS-NLMS $L = 256$ and for every WRVSS-NLMS $L = 64$.

LMS Parameters

As stated, the length L of the LMS algorithm is highly affected by the warping. This length defines the length of the resulting filter and also affects the convergence speed of the algorithm: the bigger L is, the slower the algorithm becomes. Therefore, smaller L is preferred. However, big L might be needed to sufficiently model a system. This conflict is solved by the frequency warping, which improves especially the low frequency behavior of an LMS algorithm with small L. The effect of warping can be seen in Fig. 19.

Figure 19 shows the RVSS-NLMS and WRVSS-NLMS algorithms tested with white noise (Fig. 19(a)) and pinkish bus noise (Fig. 19(b)). The lengths of the used signal samples were 22050 samples, or 0.5 seconds, and they were selected from a longer original signals of 10 seconds. In addition to warping, the two algorithms also differ in their length: in both cases the L for normal RVSS-NLMS is 256, whereas for the WRVSS-NLMS it is 64. In the topmost figures, the results from both algorithms are compared to a target value in the limited frequency band of the proposed algorithm. This target value is obtained from the same recorded signals that are used with the both LMS algorithms with deconvolution: while the LMS algorithms only use a half second segment of the recordings, the target is obtained by deconvoluting multiple small segment from the signal and then averaging the results. As can be seen from the topmost figures, both versions of the LMS algorithm produce results that represent the target curves well. Even though the WRVSS-NLMS uses lower length, it actually gives better correspondence than the longer RVSS-NLMS in major part of the frequency band especially when bus noise is used.

The middle figures in Fig. 19 show the adaptation errors in decibels for both algorithm variations. In Fig. 19(a) it can be seen that the error of WRVSS-NLMS (red curve) reaches its steady-state level more quickly than the blue curve of RVSS-NLMS error, i.e. the WRVSS-NLMS algorithm reaches its result before the RVSS-NLMS algorithm does, after which it only tracks changes. However, in the same figure it can be seen that the blue curve settles for a lower level, which means that the corresponding isolation estimation should be better and correspond the actual isolation better than the one obtained from the warped algorithm. The same phenomenon can be seen also in Fig. 19(b), although more subtly. Still, though, the curves in the both topmost figure are highly similar and shown no such substantial differences. The reason for this is the limited frequency band which is used for this analysis. Thus, the weighting discussed in Section 3.6 should be utilized so that the two variations of the LMS algorithm can be better compared in the context of this thesis work.

The errors in the middle figure are unprocessed and therefore they contain the information of the whole frequency band. This is problematic for two reasons. First of all, as mentioned in Sections 3.5 and 3.6, the frequency resolution of the human auditory system is nonlinear and as a result, estimation errors of the same magnitude at different frequencies may not be perceived similarly. Secondly, the interesting frequency band of the proposed algorithm is limited to 100–10000 Hz, whereupon analysis of the unprocessed error signal produces unrepresentative results. Fortunately, both problems can be solved by using the weighting discussed in Section 3.6, i.e. filtering the errors signals with the ERB weighting filter. The response of the filter can be seen in Fig. 16. It incorporates both the human auditory resolution and the limited interesting frequency band to the errors. These weighted error signals can be seen in the bottom plots in Fig. 19. In both Fig. 19(a) and 19(b) the errors produced by the RVSS-NLMS (blue curve) and WRVSS-NLMS (red curve) algorithms are almost equal, i.e. in the sense of human hearing resolution the errors are equally significant. Therefore, the WRVSS-NLMS with $L = 64$ produces as good results as the RVSS-NLMS with $L = 256$, but converges quicker. For these reasons the WRVSS-NLMS with $L = 64$ is chosen.

After the length of the LMS algorithm is chosen, only the κ parameter needs to be selected. It defines the memory factor α as per Equation (33). Even though Vega *et al.* used $\kappa = 1$ and $\kappa = 3$ for different kinds of signals in their algorithm comparison [88], in this work a value of $\kappa = 6$ is used. In a preliminary test it was found that this value produces the best results when used in the proposed algorithm with multiple types of excitation signals. Since the algorithm is intended to be used in everyday life and thus in many different environments, the robustness provided by the selected κ value is a great feature.

Finally, the WRVSS-NLMS algorithm produces a vector $\mathbf{w}$ as a result. It contains the warped filter coefficients estimating the impulse response of the momentary isolation of the earpiece in question. After the actual LMS algorithm, the resulted vector is averaged with four previous results and used as initial values for the next cycle of the WRVSS-NLMS algorithm. If the conditions stay constant between the time instances when the isolation is estimated, this results in almost instantaneous convergence of the RVSS-NLMS algorithm since the initial values are the correct convergence values or close to them. The actual vector $\mathbf{w}$ is then passed for the next block.

4.2.5 Dewarping

The output $\mathbf{w}$ of the WLMS block in Fig. 17 is frequency warped, and thus it needs to be dewarped, i.e. transformed from the warped frequency scale to the traditional frequency scale. This operation is conducted in the dewarp block. Dewarping a signal is closely related to warping the same signal. A signal is warped using an allpass filter chain, where every allpass filter has the same parameter λ. In dewarping a similar allpass filter chain can be utilized, but parameter is changed to $-\lambda$ instead. In the proposed algorithm the output of the dewarp block is the impulse response of the earpiece isolation h corresponding to linear frequency scale, which then can be transformed into frequency domain in the next block.

4.2.6 FFT and Smoothing

The final blocks of the adaptive isolation estimation in Fig. 17 are the FFT and smoothing blocks. First the dewarped filter coefficient vector h is transformed from the time domain into the frequency domain with FFT, which results in vector H. Third octave smoothing is then utilized to obtain momentary outcome. When this momentary outcome is used together with four previous outcomes to average the results, the final isolation levels $\hat{H}$ are obtained. However, if the disturbance detector is activated, the tables that contain the previous outcomes are reset to zero and, thus, the averaging is started anew. Together the third octave smoothing and averaging steady the isolation estimation results.

The final result from the smoothing block is the vector $\hat{H}$ containing the isolation levels. This is also the final result of the adaptive isolation estimation part of the proposed algorithm that can now be plotted and passed to the hear-through equalization part of the algorithm for EQ adaptation.

4.3 Hear-through Equalization

Hear-through equalization part of the proposed algorithm can be seen in the bottom part of Fig. 17. It adds the ARA functionality to the algorithm, i.e. it produces the hear-through signal that turns the prototype headset acoustically transparent. The purpose of this part is to run constantly and in real-time process the sound environment of the users as they interact with their surroundings so that the headset does not color the perceived sound. This is achieved with two different equalization blocks that are discussed in the next sections.

4.3.1 Microphone Inverse Equalization

Similarly to the adaptive isolation estimation part, also the hear-through equalization uses one of the earpiece microphones. The ambient sounds are recorded with the out microphones and these recordings are then processed before being played back with the earpiece transducers. The first processing block is the microphone inverse equalization.

The microphone inverse equalization is needed since the frequency response of the out microphone is not flat. The spectrum of the recorded out signal is not supposed to be white, even if white noise is used as an excitation signal. As stated in Section 2.5, when the earpiece is placed in the ear and the out microphone is thus located near the ear canal opening, spatial cues are captured in the recording. These result from reflections from the pinna and upper body of the user. However, even when these spectral alterations are taken into account, large boost can be seen between 10–20 kHz in the out signals. This is shown in Fig. 20.

The cause of the boost is not the microphone alone, but the combined effect of the

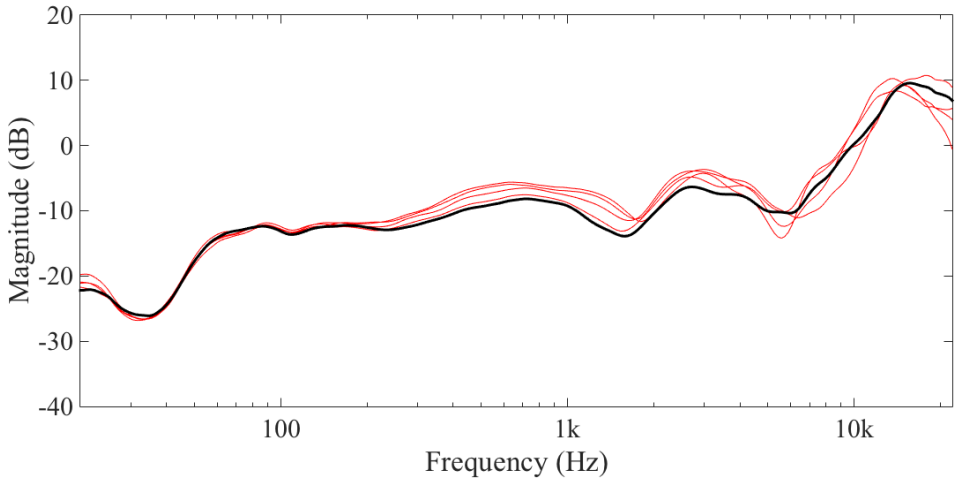

Figure 20: The spectra of out microphone recordings. Red curves are different recordings and the black curve is their average. The spectral cues from the ear and upper body reflections can be seen especially around 1–3 kHz. The emphasis of 10–20 kHz can clearly be seen as a boost of around 20 dB.

microphone, its casing, and its attachment to the mechanics of the earpiece. Since the out microphone cannot be placed directly to the surface of the casing but needs to be protected, the combined acoustics can produce resonances, which deteriorate the performance of the microphone. Additionally, resonances can occur since the microphone is attached to a circuit board.

When the boost is left unprocessed, hear-through signals are tinny and sibilant, and these problems are easily noticeable. Even though the interesting frequency band of the proposed algorithm is limited to 100–10000 Hz, an additional equalizer must be inserted to the signal chain. The equalizer is realized as a second order parametric Regalia-Mitra equalizer with constant values. The parameters of the equalizer were selected by averaging a wide array of different recordings, and they can be found in Table 1. In an informal listening test the inverse equalizer was found to improve the timbre of the hear-though signal in the desired frequency region without much affecting other frequencies.

Table 1: The parameters of the microphone inverse equalizer.

G (dB)	f_c (kHz)	BW (kHz)
-16	16	2

4.3.2 Equalization of Captured Ambient Signal

As stated in Section 2, wearing in-ear headphones affects the acoustics of the ear. This can be seen in Fig. 21. It shows three frequency responses that have been recorded with a dummy head. The black curve represents the situation, where the dummy head is not wearing headphones, i.e. its ears are open. The blue curve shows what happens to the sound reaching the eardrum, when the prototype headphones are inserted in the ears. Finally, the red curve depicts a situation, where hear-through system is used without any equalization, i.e. the recorded out signal is played back without processing. The missing peak at 3 kHz and the new peak at 6 kHz are the clearest signs of the changed acoustics. As can be seen when comparing the black and red curve in Fig. 21, hear-through equalization is required.

In the proposed algorithm, the hear-through equalization is performed with Regalia-Mitra parametric equalizers. Unlike in the microphone inverse equalization, the parameters of these equalizers are not constant, but are adjusted by the results of the adaptive isolation estimation part of the algorithm and by the user. The default settings of the hear-through equalization filters are not initialized to zero, but instead they are chosen to correspond to the required hear-through equalization in ideal conditions, i.e. when the headset is properly inserted into the ear canal. Thus, the transfer function from the headset to the eardrum must be flattened and the free air response from the out microphone to the eardrum must be duplicated [105].

The required equalization can be obtained by using signals similar to the ones shown in Fig. 21. Since the red curve depicts a situation, where the hear-through signal was not equalized, it takes into account the characteristics of the headphone

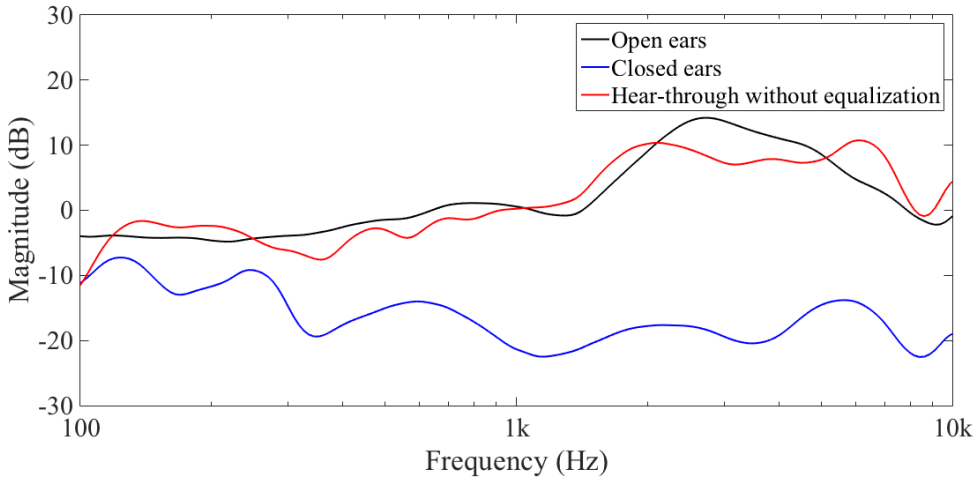

Figure 21: The effect of the prototype headset on the acoustics of the ear of a dummy head shown in the selected frequency scale of the proposed algorithm.

transducers and the out microphone. Therefore, if equalization is added to the recorded out signal, it can be seen as a change in the red curve. Since the target response is the black curve in Fig. 21, the required equalization curve can be defined by comparing the differences in the red and black curves. The required equalization can be seen as the blue curve in Fig. 22

The signals that are used to derive the required equalization curve are measured in an anechoic chamber from single directions. The used directions were between $-30°$ and $30°$ in the front sector of the dummy head ($0°$ is the direction of the nose). The use of discrete direction in the calibration of the required equalization curve results in a correct hear-though experience when the headset is used to listen to sounds

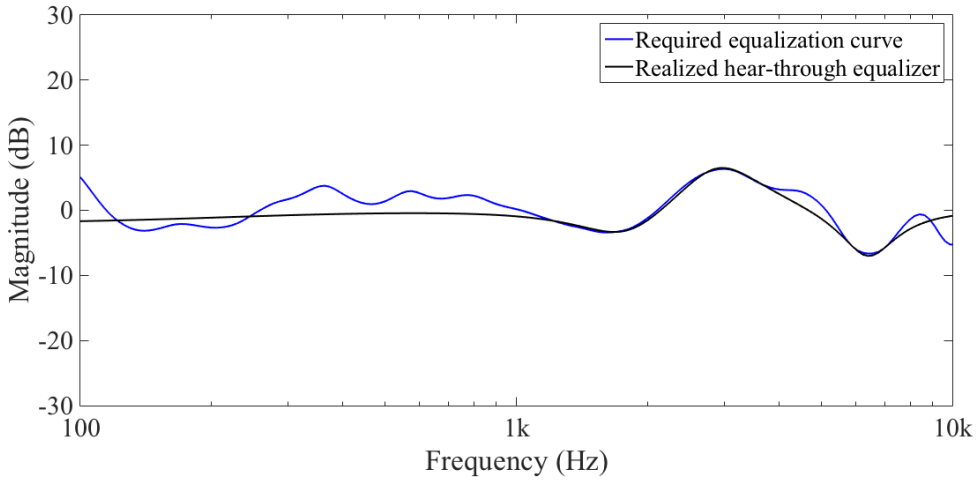

Figure 22: The required equalization curve and the realized hear-through equalizer compared with each other.

coming from the front sector of the user, but may result in slightly colored timbre for sounds from other directions. This is an informed compromise, since this specific calibration was chosen to maximize the naturalness of important sounds. When people concentrate on an event or other person that that they consider important, they usually turn to face that target of concentration. For example, when talking with another person, watching television, or attending a rock concert the main direction for coming sounds is the front sector, and thus the proposed algorithm produces a correct hear-through signal for acoustical transparency.

To realize the required equalization curve, four parametric equalizers are used. One of those is a first-order shelving filter and the other three are second-order peaking filters. The shelving filter is needed since the isolation of the prototype headset is weakest at low frequencies and, thus, the low frequency components leak the most to the ear canal. The second and third peaking filters are, on the other hand, utilized to counter the changes in the acoustics of the ear caused by the headset. As stated in Section 2.3, in-ear headphones cancel the natural resonance around 3 kHz and create a new, unnatural one around 8 kHz, and therefore the equalizers need to work vice versa, i.e. to create the resonance around 3 kHz anew and to cancel the resonance around 8 kHz. The default parameters of the parametric equalizers can be seen in Table 2 and the corresponding frequency response as the black curve in Fig. 22.

As can be seen in Fig. 22, the realized equalizer does not follow the blue curve strictly below 1 kHz. In order to fit the equalizer better, two more second-order parametric equalizer sections would be needed, thus making the equalizer controls more complex. Additionally, when the complex equalizer was compared with the realized equalizer in an informal listening test, it was found that the added complexity did not have a clear effect on the timbre of the hear-through signal, and therefore the simpler construct was chosen.

Table 2: The default parameters of the hear-through equalizers.

	G (dB)	f_c (kHz)	BW (kHz)
Shelf	-2	0.2	-
Peak 1	-6	1.8	0.9
Peak 2	9	2.9	1.2
Peak 3	-8	6.4	1.6

Automatic Hear-Through Equalization Control

Now that the derivation of the default equalization parameters is complete, it is time to introduce the parameter controls, first of which is automatic. Even though the two parts of the proposed algorithm are presented here separately, they are linked together as can be seen from the block diagram in Fig. 17: the isolation level information is used to control the default equalization settings. Since the default settings are optimized for situations when the headphones are properly fitted, i.e. the

isolation is approximately at its largest value, some adjustments are needed when the isolation levels decrease. Otherwise the combined effect of the leaked sound and the hear-through signal may result in either a boosted response when compared to the target open ear response, or to a comb-filtering effect [15].

When the headphones are inserted poorly to the ear canal, two things occur: the decreased isolation leads to increased leakage of surrounding sounds especially at low frequencies and the frequency response of the headset is deteriorated at low frequencies due to the pressure chamber principle (see Fig. 4). The deteriorating frequency response is actually advantageous since it automatically results in fewer low frequency components in the hear-through signal, and thus the increased leakage may not result in the comb-filtering effect. However, the response may still be boosted around 1 kHz since those frequencies are not affected by the lack of the pressure chamber principle. in order to ensure the best possible acoustical transparency, the low frequency response of the hear-through equalization is attenuated up to 2 kHz according to the deteriorated isolation.

The low frequency behavior of the hear-through equalization can be controlled with one of the Regalia-Mitra equalizers, namely the shelving filter. By default, its effect on the equalization is minor, but by adjusting the gain and the crossover frequency the whole interesting frequency band can be affected. However, since it was found that the poor fit of the headphones mainly affects the response at 2 kHz and below, the adjustments are limited correspondingly. The required gain adjustments are derived from the isolation level value at 1 kHz: the current value is compared to the optimal value measured with the dummy head, i.e. approximately 20 dB, and when the isolation has deteriorated 5 dB or more, the gain of the shelving filter is adjusted. The adjustments are done every time the estimated isolation decreases 1 dB with nonlinear steps, which are chosen to result in 1 dB change in the equalization response at 1 kHz. There are ten steps defined, and therefore the gain remains constant after he isolation has decreased to 6 dB. This result in maximum gain adjustment of 22 dB and maximum attenuation change of approximately 10 dB at 1 kHz, which corresponds to the maximum excessive boost caused by a poor fit. Finally, since the crossover frequency of the shelving filter remains the same during the automatic gain adjustments, the gain of the second peaking filter is compensated in order to limit the effect of the shelving filter above 2 kHz.

Manual Hear-Through Equalization Controls

In addition to the default equalization settings and the automatic setting control, the proposed algorithm also offers user controls. These are represented by the EQ control and volume control blocks. The user controls offer possible improvements in the hear-through signal and thus more suitable equalization for certain situations or sound environments. For example, the speech intelligibility may be increased during a hear-through event by boosting the equalization around 3 kHz, or a piercingly bright sound may be equalized to be more pleasant by adjusting the gain of the high frequencies.

The EQ controls offer three sliders that can be used to adjust the gain of low,

middle and high frequencies. These affect the same equalizers that form the default hear-through equalization. The sliders range from −10 dB to 10 dB in 1 dB steps, and therefore they offer changes that are perceivable, but not so great to add clipping with normal signal levels.

The low frequency slider affects directly the gain of the shelving filter. Since the gain adjustments are small and the crossover frequency of the shelving filter is low, the gain adjustments do not cause the response of the shelving filter to leak to the frequency bands of the other equalizers. Therefore, the crossover frequency can remain constant. However, if the isolation level estimates have caused adjustments in the shelving filter parameters, the bass slider is disabled and reset to zero until the optimal fit and, thus, the optimal isolation is achieved again. This minimizes the risk of comb-filtering effect caused accidentally by the user.

The middle EQ control slider is used to adjust the equalization gain in the middle frequencies. It affects both the first and second peaking filter in order to offer wide band control around 2–4 kHz. However, the center frequencies and the bandwidths of the two peaking filters stay constant. The slider setting affects the gain of the first peaking filter by the amount of $\frac{g_{mid}}{3}$, where g_{mid} is the value of the slider, and the gain of the second peaking filter by the amount of $\frac{2g_{mid}}{3}$. Additionally, the gain of the second peaking filter is corrected by an offset amount, which depends on the position of the third slider: the offset is zero when the third slider value is greater or equal to zero, but when the slider value is less than zero, the offset is added. It takes into account the combined effect of all the equalizers at the center frequency of the second peaking filter. The offset value is calculated as the change of the combined effect when compared to the default equalizer settings.

The final EQ control slider, i.e. the high slider, adjusts the gain of the third peaking filter. However, similar to the middle slider, an offset value is also utilized here. It takes into account the combined effect of the second and third peaking filter by correcting the new gain of the third peaking filter. The offset is equal to the change in leakage at the center frequency of the third peaking filter that is caused by the possible changes in the gain of the second peaking filter. Finally, also with third peaking filter the center frequency and the filter bandwidth stay constant regardless of the changes in gain.

Manual Transparency Adjustment

One final user-adjustable slider that controls the equalization is the volume control, which ranges from −30 dB to 10 dB. It gives the users a possibility to slightly boost the hear-through signal, if they want perceive a sound environment louder than they would normally without an ARA headset. However, the boost is limited to 10 dB so that the combined maximum amplification of all the equalization sliders and the volume slider is less likely to result in clipping. The highest peak in the equalization curve is approximately 20 dB when all the sliders are at maximum position. On the other hand, the more extensive attenuation range offers a possibility to perceive the sound environment quieter and thus utilize the passive attenuation of the headset.

When using the volume slider to boost the hear-through signal, the hear-through

equalizer is linearly scaled on the dB scale. On the other hand, when the slider is used for attenuation, the equalizer is scaled downwards and in addition the shelving filter is adjusted. The response of the shelving filter must decrease in bigger steps than the rest of the equalizer to reduce the comb-filtering effect. Since the isolation of the prototype headset is weakest at low frequencies, those frequencies leak to the ear canal. When the amplitude of the hear-through signal is decreased, the leaked sound remains the same, and when the amplitudes of these two signals approach each other, the comb-filtering would become more and more noticeable. However, if the low frequencies are attenuated from the hear-through signal, the leaked sound is predominant and the comb-filtering effect is reduced.

The reduced low frequency response is achieved by adjusting the gain of the shelving filter while also reducing its crossover frequency by $5g_{vol}$ Hz, where g_{vol} is the value of the volume slider. The gain adjustment steps, on the other hand, must be nonlinear and range from 0 dB to -44 dB while the volume slider is adjusted between 0 and -30 dB. The calibration point for the attenuation was selected at 3 kHz so that one volume slider step would result in approximately 1 dB change in the equalizer response at 3 kHz. Since the comb-filtering effect starts to increase when the hear-through signal is attenuated 5 dB or more, the gain adjustment steps are initially large, i.e. 2–3 dB for one volume slider step. However, the steps are quickly decreased to approximately 1 dB to minimize the effect of the shelving filter above 1 kHz, since the comb-filtering effect is only noticeable at lower frequencies. Finally, to further minimize the comb-filtering effect, the bass slider is disabled and its value reset to zero, when the volume slider is used to attenuate the hear-through signal. Thus, the users themselves cannot accidentally increase the comb-filtering.

The output of the equalization of block in Fig. 17 is the hear-through signal. At the same time it is also the end result of the whole proposed algorithm. The obtained hear-through signal is next fed to the transducers in the earpieces for playback, which results in acoustical transparency with the default settings or to a possibly enhanced perception of surrounding sounds, when the equalization controls are used.

4.4 Matlab Implementation

This section covers the Matlab implementation of the algorithm, which also utilizes Playrec. Playrec is a free cross-platform Matlab audio utility that offers buffer-based processing of signals, and for example simultaneous playing and recording [106]. The Matlab implementation of the proposed algorithm can be used to demonstrate its operations in pseudo real-time. For the demonstration, a following setup is needed: a computer to run the Matlab simulator, a sound card with at least four input channels for all the headset microphones and a headphone stereo output, and an external loudspeaker that is used to produce the ambient sounds.

The adaptive isolation estimation part of the proposed algorithm is implemented according to the block diagram in Fig. 17. First, the external loudspeaker is used to simulate a noisy outdoor environment. The recorded signals from out and in microphones are then processed as described in Section 4.2, which results in individual isolation estimations for both headset earpieces. However, since the Matlab

implementation does not allow actual real-time processing, Playrec is used with a buffer size of one second. Therefore, the resulting isolation levels describe the isolation as it was during the previous buffer window, while the algorithm is calculating the current isolation estimates at the same time. In addition the WLMS algorithm is only provided with half of the buffer window data to reduce the computational costs. Regardless of this, the proposed algorithm is able to both converge to the solution and track the changes occurring during the half second windows. Finally, dividing the WLMS algorithm in a such manner instead of running it constantly also improves the tracking ability of the RVSS-NLMS algorithm. Since the δ-sequence converges to zero, the tracking ability of the algorithm is lost. But when the algorithm run separately for each buffer window, the δ-sequence is calculated anew for each window, which enables tracking.

The hear-through equalization part of the proposed algorithm, on the other hand, cannot be implemented in Matlab as it is shown in Fig. 17, but instead the block diagram of the implementation can be seen in Fig. 23. This implementation simulates real-time use of the proposed algorithm with the help of external loudspeaker and the headset. Additionally, the two parts of the Matlab implementation cannot be performed simultaneously. Instead, first the Adaptive isolation estimation is performed, which results in estimated isolation levels. These results are then utilized to adjust the equalization settings. Also, the possible equalization adjustments done by the user are taken into account, and finally the hear-through part is demonstrated. The differing functionalities between the proposed algorithm and the Matlab implementation include the signal selection, the delay adjustment and the playback of the signals, and they will be discussed next.

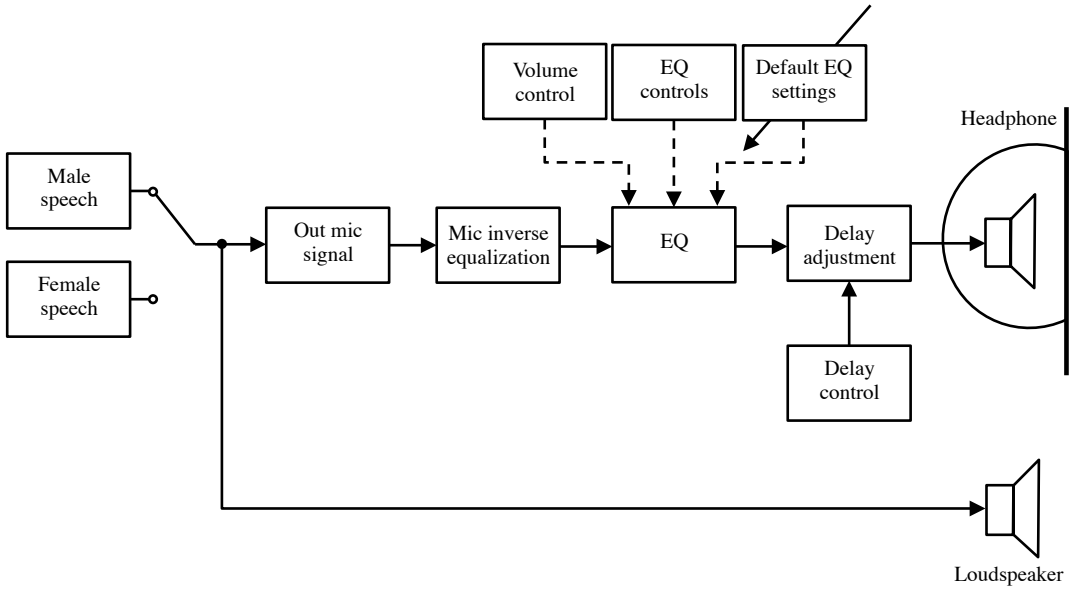

Figure 23: Block diagram of the Matlab implementation version of the hear-through equalization part of the proposed algorithm.

4.4.1 Signal Selection and Playback

The purpose of the hear-through equalization part of the proposed algorithm is to remove the effects of the headset earpieces on the acoustics of the ear, i.e. to transform the headset acoustically transparent. This is achieved by playing an equalized version of the surrounding sounds to the user with the headset at the same time they occur, or with minimal delay. Since the Matlab implementation of the proposed algorithm is not running in real-time, the hear-through equalization must be simulated. The simulation is achieved with the help of controlled ambient sounds and an external loudspeaker.

First, the users must select the excitation signal. They can choose between a speech signal from either a male or a female speaker. The spoken sentences are same with both speakers. The chosen signal is then reproduced with the external loudspeaker and recorded with the out microphones of the headset. Next, the recorded signals go through the processing shown in Fig. 23. As stated before, the block diagram shows a mono system, which is applied to both signals separately. The equalization is the same for both signals.

After the signals have been processed, they are fed to the transducers of the earpieces and played back. However, the original signal is also reproduced at the same time with the loudspeaker to achieve the effect that the user is listening to hear-through signals. The delay control setting affects the exact timing of the signals with respect to each other, and it will be discussed next.

4.4.2 Delay Adjustment

Since the proposed algorithm was not realized in real-time, the comb-filtering effect cannot be evaluated properly. However, to demonstrate it, a delay adjustment was added on the Matlab implementation together with user-adjustable delay control.

First of all, the delay adjustment block removes all the latency caused by the sound card while recording or the equalization. By default, the user can observe the ideal circumstances, i.e. the hear-through signal with zero delay and thus without the comb-filtering effect. However, the delay controls offer possibility to obverse the nonideal circumstances. The delay can be adjusted between 0–10 ms, which contains the aforementioned ideal situation, a possible latency of a real-time realization (around 1–4 ms), and even a larger delay that corresponds to a nonsuitable DSP-board. When the delay slider is not zero, the hear-through signal is delayed by inserting zeros at the beginning of the signal before the playback.

4.4.3 Algorithm Implementation GUI

Figure 24 shows a screen shot of the graphical user interface (GUI) of the Matlab implementation. The two parts of the algorithm are clearly visible in the GUI and can be simulated separately. First, the simulator must be initialized. Next, adaptive isolation estimation is started by selecting “Adaptation.” The estimated isolation levels of both earpieces are plotted to the left figure: the blue curve is the isolation of the right earpiece and the red curve is the left one. At the same time the right figure

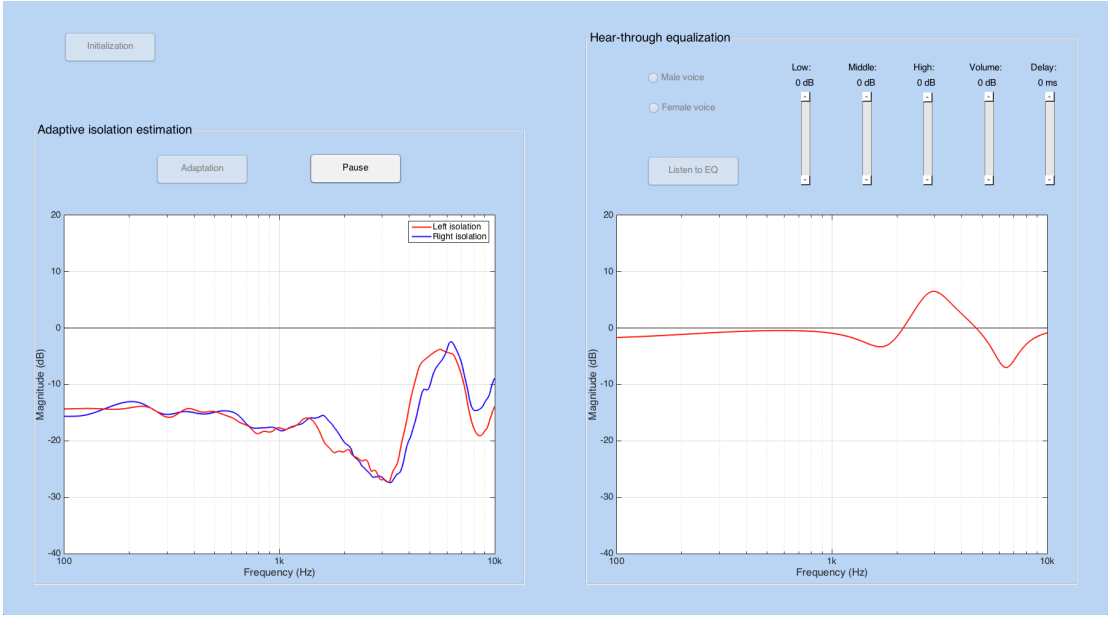

Figure 24: The GUI of the proposed algorithm Matlab simulator. The left plot shows the estimated isolations for both ears and the right plot shows the current response of the hear-through equalizer.

shows the current hear-through equalization curve. When the adaptation is paused, the equalization can be tested with the hear-through equalization part. First, the excitation signal is selected. Then the equalization curve, which has been adjusted according to the estimated isolation levels, can be further adjusted with the sliders. The low, middle and high sliders adjust the equalizer curve as described in Section 4.3.2. The volume slider can be used to alter the level of the equalization curve, and the delay slider is used to simulate the comb-filtering effect. Finally, the hear-through equalization can be listened by pressing the “Listen to EQ” button.

5 Results

In this section, the behavior of the proposed algorithm is validated with measurements and simulations. The frequency responses obtained from the different functions of the proposed algorithm are compared to target values to show, how well the processing achieves its goals. The validations are done in the interesting frequency band, i.e. 100–10000 Hz.

The structure of the section is as follows: First the measurement setup is described. Then, the isolation estimates from the proposed algorithm are compared to isolation result obtained with frequency sweeps and deconvolution. Additionally, the two differing isolation values of the headset are explained. Next, the results from the hear-through equalization part are analyzed thoroughly. This includes the equalization with default settings in ideal conditions, the equalization for sounds coming from other directions than the frontal sector that was used in the equalizer calibration, the effect of the mic inverse equalization, the effect of the automatic settings control in poor fit situations, and, finally, the effect of the user controls.

5.1 Measurement Setup

The validation measurements were done mainly in an anechoic chamber in Aalto University School of Electrical Engineering. The chamber in question can be seen in Fig. 25 together with some of the test equipment. In addition, some of the isolation measurements were performed in an ordinary office room in order to obtain results from a real-world environment instead of the ideal environment of the anechoic chamber.

Figure 26 shows the dummy head that was used in the validation measurements. It is Head Acoustics HMS II.3-33 artificial head measurement system that is designed according to the requirements of ITU-T Recommendations P.57 and P.58 [107]. It has anatomically shaped pinna simulators, and therefore it is suitable for intra-concha

Figure 25: The anechoic chamber used in most of the validation measurements.

Figure 26: The Head Acoustics dummy head that was extensively used to measure the performance of the prototype headset and the proposed algorithm.

headphone measurements. The microphones of the dummy head are located at drum reference point (DRP), i.e. inside the ear canal at the position, where the ear drum of a human would be.

The signals from the dummy head were then fed through a MOTU UltraLite-mk3 Hybrid sound card [108], which was connected to a Macbook laptop. Matlab was used to create the measurement signals and analyze the results. Finally, the external excitation signals were played back with a Genelec 8030B loudspeaker [109].

5.2 Isolation Estimation Responses

The first part of the proposed algorithm is the adaptive isolation estimation. Its output is the vector containing the estimated isolation levels, and if this output is plotted, it can be compared with isolation results obtained with other techniques.

In order to validate the algorithm result, three isolation vectors are calculated: one using the dummy head microphones and two using the headset microphones. The complete isolation can be obtained by playing two logarithmic frequency sweep and recording them with dummy head. For the first recording the ears of the dummy head are open and for the second the ears are blocked by the headset. This results in two impulse responses, which can be deconvoluted by FFT to obtain isolation impulse response. This can be seen in Fig. 27 as the black curve.

The second isolation vector is an estimate obtained with the headset microphones while using a sine sweep as an excitation signal. The sweep is played once and recorded with the headset microphones that are inserted to the ears of the dummy head. Similarly to the first isolation vector, the isolation estimate impulse response is obtained by deconvolution. The result can bee seen in Fig. 27 as the blue curve. This isolation estimate shows the ideal result that can be achieved with the headset

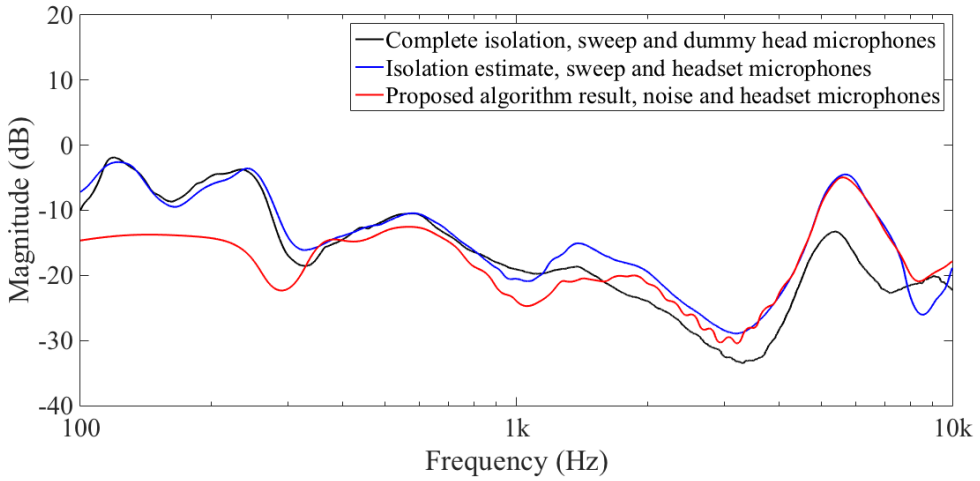

Figure 27: The isolation estimate from the proposed algorithm compared with the complete isolation and ideal isolation estimate.

microphones.

Finally, the third isolation estimate vector is obtained with the proposed algorithm. First, a short noise signal is played. The noise used for this test is pinkish noise that simulates traffic noise. The noise is captured with the headset microphones and fed to the adaptive isolation estimation part of the proposed algorithm. The result can be seen as the red curve in Fig. 27.

There are differences between the isolation vectors in Fig. 27. First of all, the differences between the blue and black curve can be explained with different measurement locations and the mechanics of the headset. Even though both of the aforementioned curves represent ideal situations, the first is measured at DRP, whereas the other is measured at the beginning of the ear canal. Thus, the response of the ear canal is not present in the isolation estimate, i.e. the blue curve. Additionally, the earpieces of the headset allow some frequencies to propagate through them without attenuation. As can be seen in Fig. 27, the blue and red curve, both of which are recorded with the headset microphones, show a attenuation close to 0 dB between 5 and 6 kHz. This means that at those frequencies, the magnitude of the frequency components is almost equal in both out and in microphones. Therefore, when analyzing the isolation estimates that are obtained with the headset microphones, it must be remembered that they differ from the complete isolation.

Figure 27 also shows that the isolation estimate of the proposed algorithm differs from the ideal isolation estimate. The biggest difference can be seen below 300 Hz, since the proposed algorithm preprocesses the recorded signal. For example, the dc blocker removes low frequency components from the signal and its attenuation extends to approximately 300 Hz. Another frequency band, where differences can be found, is between 1–2 kHz. These differences could be more significant, since the isolation estimate value at 1 kHz is used for the equalization control. Therefore gross errors could lead to completely wrong adjustments. However, despite the small offset at 1 kHz when compared to the ideal isolation estimate, the proposed algorithm

produces stable and linear result, which are suitable for the equalization control. All in all, the red curve resembles the shape of the blue curve well, especially when one takes into account that the latter is an ideal solution, whereas the former is obtained with a practical algorithm, which is designed more for robustness than for infinite precision.

5.3 Hear-through Equalization Responses

In this section, the behavior of the second part of the proposed algorithm is validated, namely the hear-through equalization part. The results presented in this section were measured by simulating real-time operations, since the algorithm was not realized with a DSP-board. The test signals were white noise that was reproduced with the Genelec loudspeaker and recorded with the headset microphones. After the equalization and other processing had been performed, the recorded signal had become the hear-through signal. It was then reproduced with the headset transducers, while the original noise was simultaneously played back again with the loudspeaker. This way the resulting sound inside the ear canal of the dummy head resembles the hear-through event. When it was recorded with dummy head microphones, it could be compared to the target open ear response recorded with the dummy head using the original white noise.

5.3.1 Default Hear-through Equalization Settings

Figure 28 shows the effect of the hear-through equalization using the default settings with different excitation signal levels. As can be seen, the blue curves follow the shape of the red curves closely at all levels, i.e. the hear-through event produces a similar frequency response to the target open ear response. Therefore, the default equalization works well, although ideally the red curves and corresponding blue curves should be equal. Furthermore, it can be stated that the out microphone of the headset works linearly at least at the magnitude range used in the measurement: since the equalization stays constant, the only variable is the magnitude of the excitation signal, which must cause a similar magnitude change in the out microphone, or otherwise the blue curve would not follow the red ones as precisely. This linearity enables the use of constant equalizer settings in ideal situations.

5.3.2 Microphone Inverse Equalization

One of the blocks in the hear-through equalization is the microphone inverse equalization. Its function is to correct the boost in the microphone frequency response by using a constant equalizer filter, which creates a notch in the overall equalization frequency response. The effect of the microphone inverse equalization can be seen in Fig. 29. The red curve shows the target response and the other two the recorded hear-through events. As can be seen, the blue and black curve are approximately equal up to 5 kHz, after which they differ. The black curve is recorded without the microphone inverse equalizer, and thus its response is boosted between 10–20 kHz, when compared to the target. This would result in excessively bring sound, which

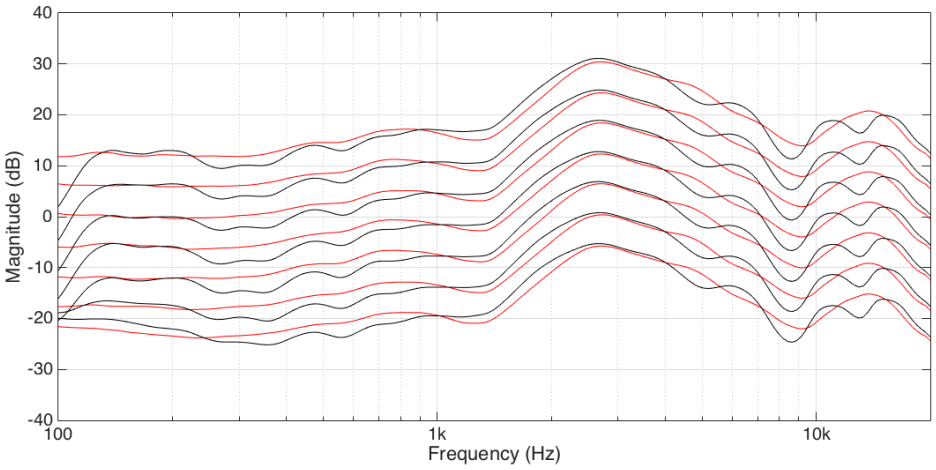

Figure 28: The effect of the default equalization settings with gradually strengthening excitation signal level. The red curves are the target open ear responses and the blue curves show the combined effect of the leaked sound and the hear-through equalized sound. The magnitude of the used excitation signal was increased by 3 dB between each measurement.

would be noticeable. However, when the microphone inverse equalizer is added to the signal chain, the resulting response in more similar to the target as can be seen by comparing the blue curve with the red one. Thus, the microphone inverse equalizer is beneficial and improves the sound quality of the proposed algorithm.

5.3.3 Other Directions

The behavior of the proposed algorithm was also tested from nonideal directions, i.e. from directions that were not used for the hear-through equalization calibration. The results from these measurements are shown in Fig. 30. As can be seen, both the black and the red curves differ between different plots. The red curves, i.e. the open ear responses, differ since the head, the pinna, and the upper body create different reflections for sounds coming from different directions. Thus, also the signal that is recorded with the out microphone of the prototype headset will depend on the angle, and since the equalization stays constant, the resulting responses, i.e. the black curves, differ from each other.

Figure 30 shows that the constant equalization settings create good match with the target curve up to 4 kHz with all the shown directions. Above that, the proposed algorithm produces attenuated responses due to the selected calibration. However, since the auditory resolution of the humans is weighted towards the low frequencies (see Section 3.6), the errors seen in Fig. 30 are not especially significant. More important is that the low frequency behavior of the proposed algorithm is precise, since errors at those frequencies would be much more noticeable.

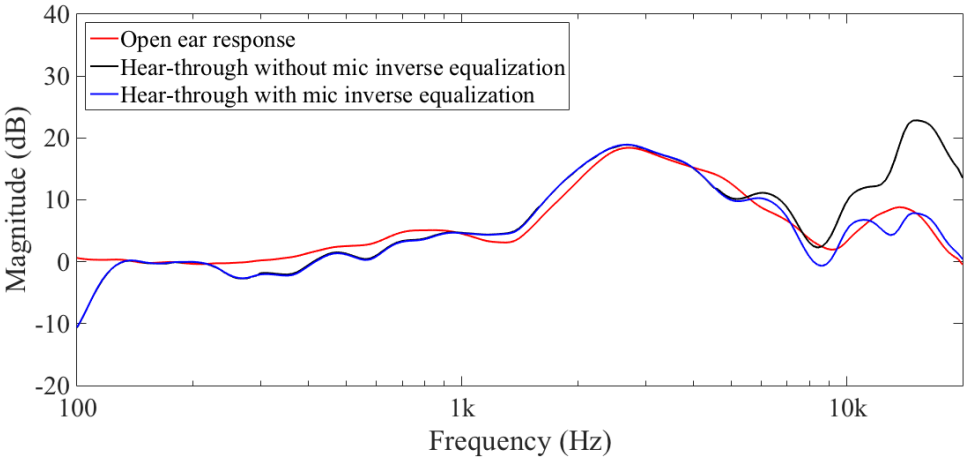

Figure 29: The effect of the microphone inverse equalization. The red curve is the target open ear response measured with the dummy head. The black and blue curves are both obtained by using the hear-through functionality of the proposed algorithm. The default equalization is used in both situations, but in addition the microphone inverse equalization was utilized when measuring the blue curve. White noise was used as the excitation signal.

5.3.4 Poor Fit Situations

In section 5.3.1 the behavior of the proposed algorithm was presented in ideal conditions. However, this is not always the case and one significant variable in the hear-through use of the headset is the fit. If the earpieces are not correctly inserted into the ear canals, the resulting hear-through events may be spectrally colored, which affects the acoustical transparency. Therefore, the equalizer response must be adjusted according to the estimated isolation.

Figure 31 shows simulated results that show the effect of the automatic equalizer adjustment. The simulations are based on actual measurements: the effect of the equalization is isolated from the recorded dummy head signals in order to adjust the equalizer and see the resulting changes in the hear-through event. This can be done, when the out microphone signal, the leaked signal, and the final hear-through event signal are known. In Fig. 31 six different poor fit scenarios are shown. In all plots, the red curve is the target open ear response, the green curve is the hear-through event with default equalizer settings, and the black curve is the simulated response of the automatically adjusted hear-through equalization.

As can be seen in Fig. 31, the adjusted equalization improves the responses in all the example scenarios and brings them closer to the target curves. The interesting frequency band in these plots is approximately 100–3000 Hz, since the poor fit affects mainly those frequencies. In narrow bands the behavior may be slightly worse (for example around 1 kHz in Fig. 31(a)), but overall the responses are closer to the target curves, and thus offer better acoustical transparency even when the earpieces

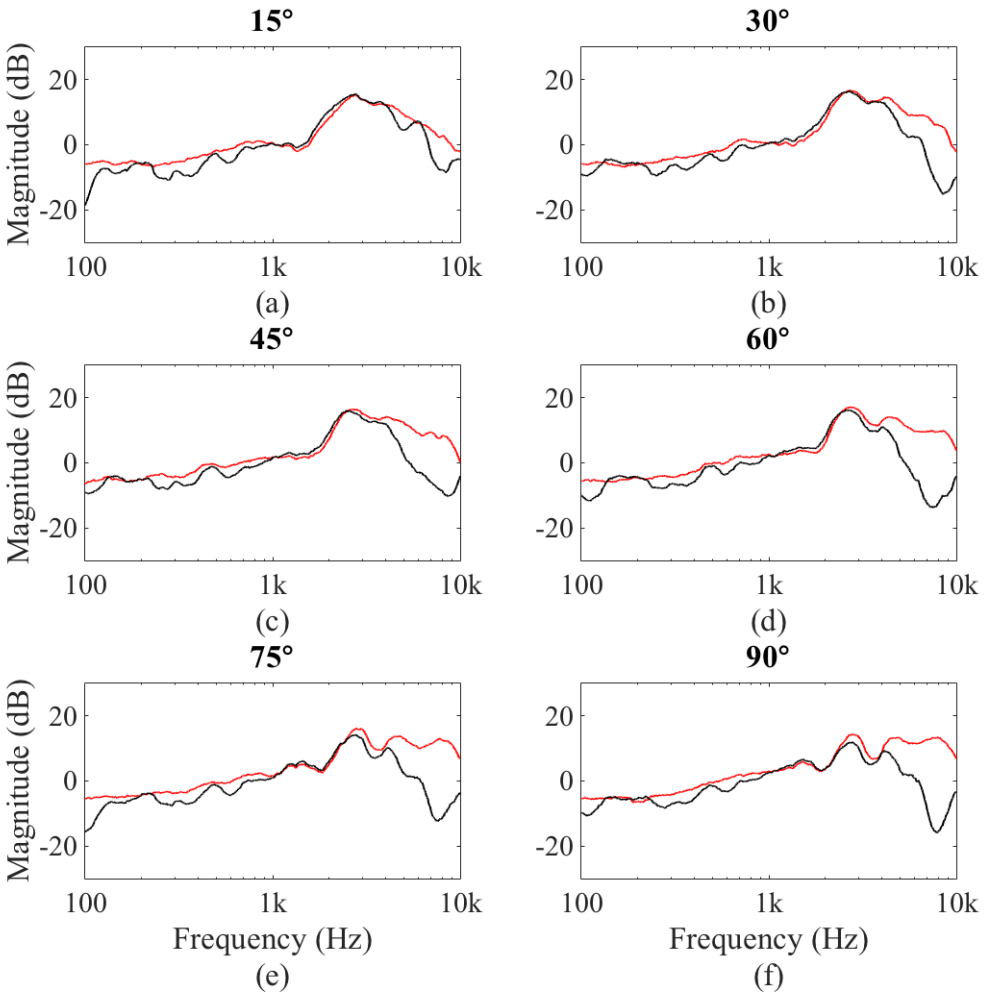

Figure 30: The behavior of the hear-through equalization measured from other directions. The red curves are the open ear target responses and the black curves are obtained by recording the hear-through events.

are not correctly inserted.

5.3.5 User Controls

The proposed algorithm provides user controls for low, middle, and high frequency behavior of the hear-through equalization, as well as the overall volume of the hear-through signal. In addition, the delay of the hear-through signal can be adjusted. The effect of EQ control sliders at their maximum setting can be seen in Fig. 32. As can be seen, the effect of the low slider is modest, since the control was kept straightforward. Therefore, large adjustments would quickly affect much broader frequency band, since the utilized shelving filter is first-order construct. Additionally, the response of the prototype headset is weakest at low frequencies. This is due to

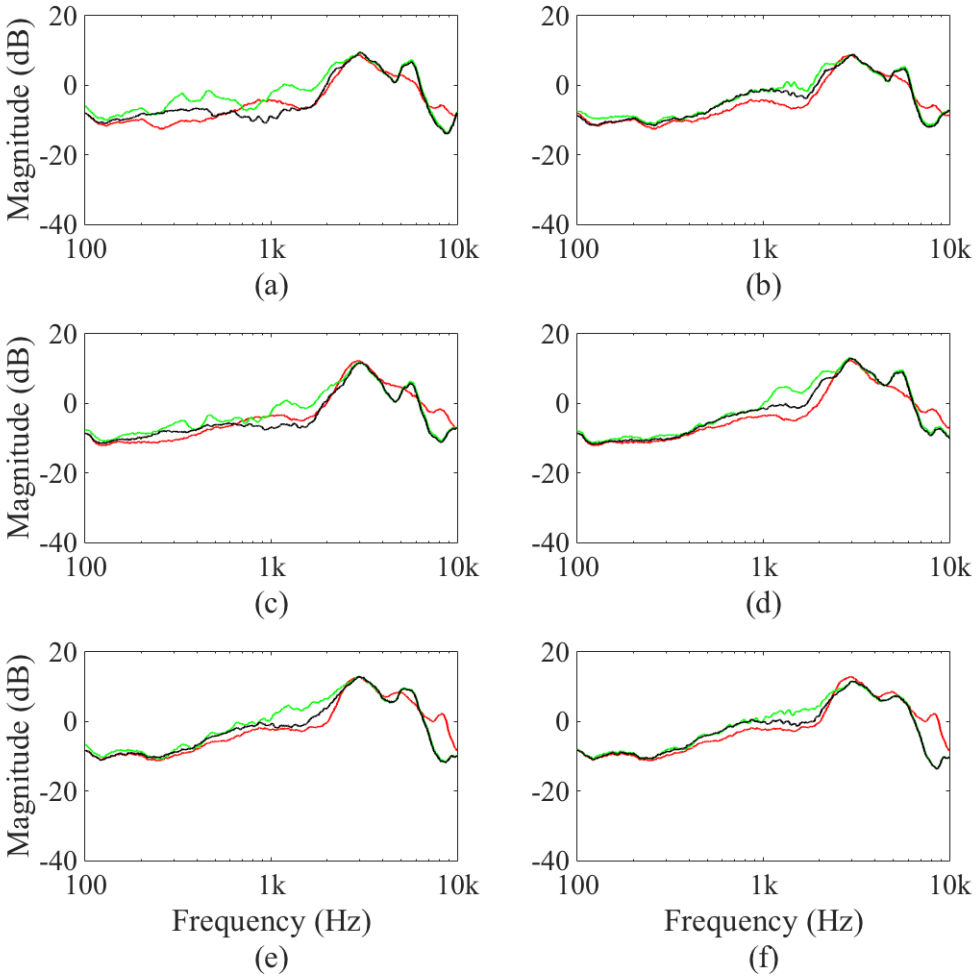

Figure 31: The effect of the automatic equalization control in case of a poor headphone fit. The plots show six different poor fit situations in which the hear-through equalizer has adapted to the decreased isolation. In every plot the red curve is the target open ear response, green curve is the simulated response with the default equalizer settings, and the black curve shows the effect of the automatically adjusted settings.

small pressure chamber effect: since the isolation of the headset is poor around 100 Hz, it can be noted that the ear canal is not blocked tightly.

Figures 32(b) and 32(c) show the effect of the two other EQ control sliders. The middle slider offers great control over the perceived hear-through sound between 1 and 5 kHz. Thus it can be used to adjust the prominent peak found in the open ear response, and since this peak is caused by the natural resonance of the ear canal, the middle slider offers extensive control over the perceived sound quality. On the other hand, the high slider affects frequencies above 4 kHz, and can thus be utilized to control the brightness and brilliance of the hear-through sound.

The effect of the volume control is shown in Fig. 33 and 34. Similarly to re-

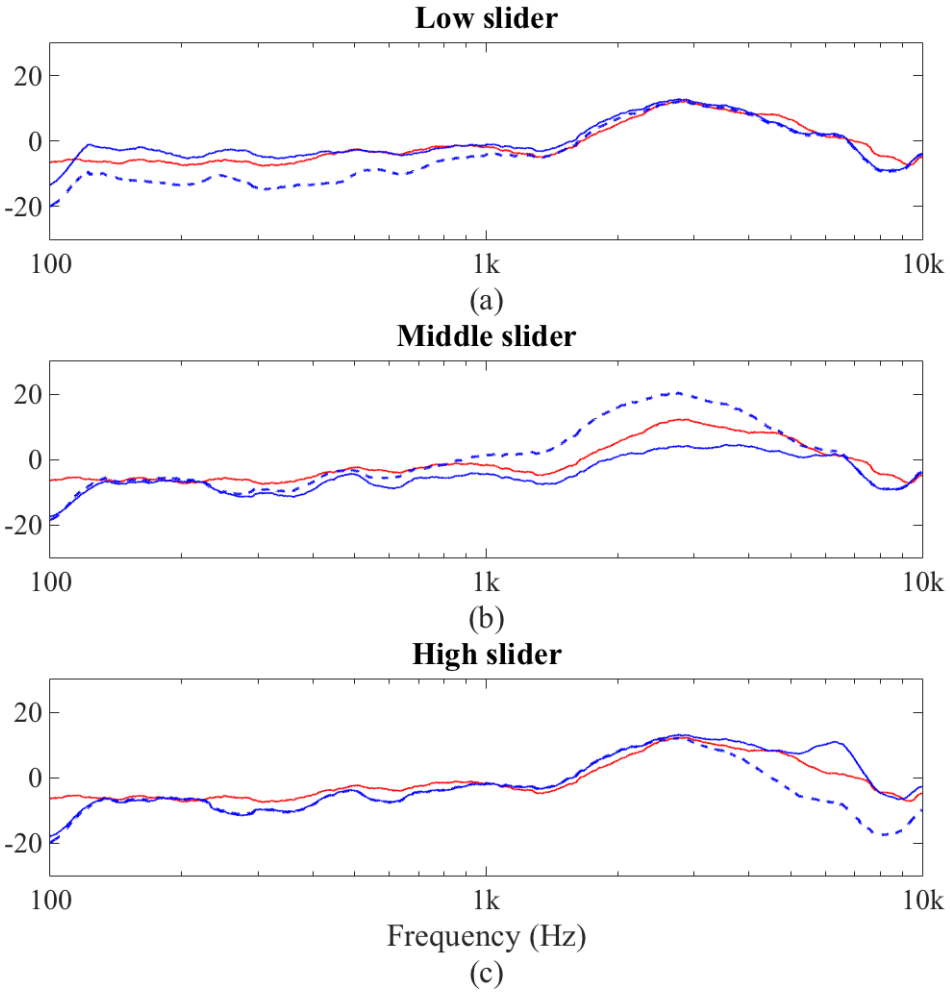

Figure 32: The effect of the user controlled sliders when set individually to maximum values. The red curves are the target open ear responses, while the solid blues curves show the effect of each slider at its maximum value and the dashed blue curves at minimum values.

sults presented in Section 5.3.4, these results are also simulated based on actual measurements. Figure 33 shows the comb-filtering effect that is caused by the lack of additional low frequency control while lowering the volume of the hear-through signal. The comb-filtering can be seen as ripple below 1 kHz. However, when the low frequency response of the hear-through equalization is attenuated quicker than the response at other frequencies, the comb-filtering effect can be avoided. Figure 34 shows the significantly improved behavior of the simulated hear-through event: when the low frequencies are removed from the hear-through signal, the ripple vanishes from the responses, since the leaked sound is predominant. Therefore, additional shelving filter control while adjusting the volume of the hear-through signal is advantageous and improves the acoustical transparency.

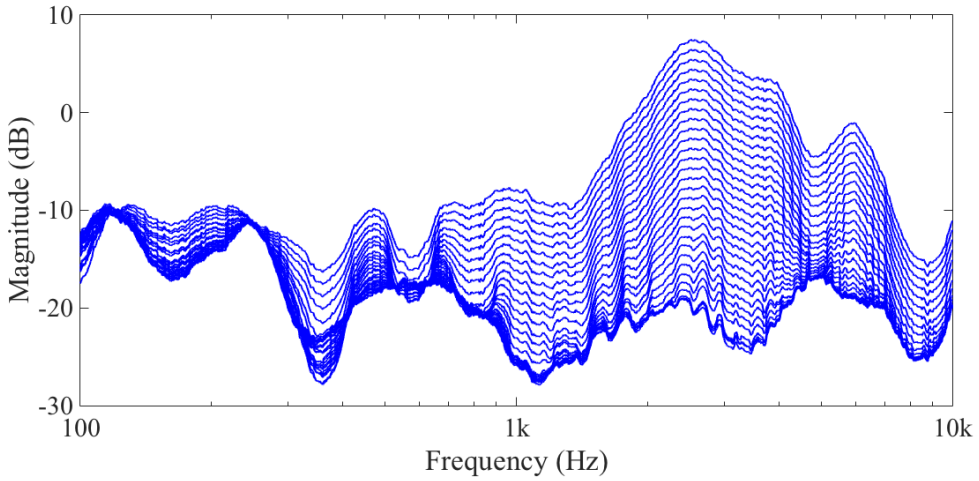

Figure 33: The effect of the volume slider without additional shelving filter adjustments. The comb-filtering effect can be seen below 1 kHz.

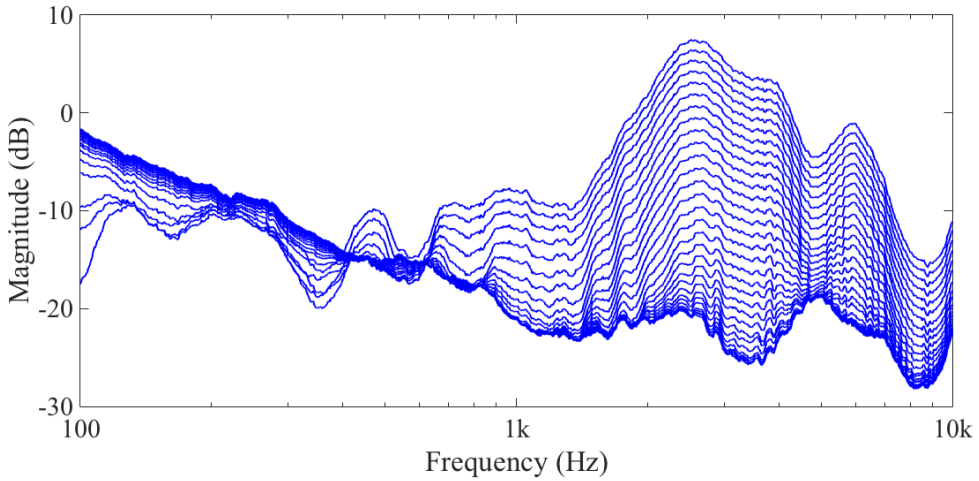

Figure 34: The effect of the additional shelving filter adjustments incorporated to the volume slider.

Finally, Fig. 35 shows the effect of the adjusted delay on the measured hear-through event. As can be seen, the delay affects the response only below 1 kHz due to the comb-filtering effect. Therefore, the differences between the target curve and the other curves are the larger the lower the frequency is.

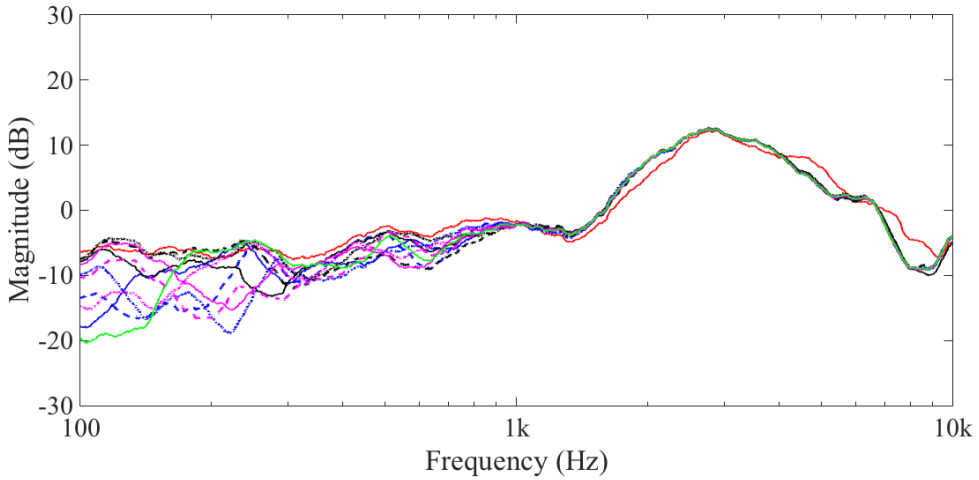

Figure 35: The effect of the user controlled delay. The red curve is the target open ear response and the other curves are hear-through events, where the hear-through signal is delayed 1–10 ms.

6 Conclusion

This thesis work presents a novel adaptive hear-through algorithm with two distinct functions: isolation estimation and hear-through equalization. The proposed algorithm utilizes a special prototype headset with both internal and external microphones to estimate the isolation properties of the headset online. The hear-through equalization is utilized to counter the effects of the headset on the acoustics of the outer ear, i.e. to achieve acoustical transparency. Furthermore, the isolation estimate can be used to adapt the hear-through equalization, which results in good acoustical transparency even with poor fitting. Finally, the algorithm also offers user controls, and thus a possibility to further improve the hear-through equalization to suit different situations.

Since headphones are nowadays used increasingly while travelling in noisy environments, the isolation characteristics of headphones have two conflicting demands: on the one hand the isolation should be sufficient to block the undesirable ambient noise from disturbing the listening, but on the other hand desirable sounds such as conversation should be audible and intelligible despite the headphones. Therefore ARA headsets such as the one used in this work have major advantages over traditional headphones that are used with mobile devices. With the help of the hear-through function, desired sounds can be let through the headphones to produce a natural sound perception, while the passive attenuation of the headset can be used to attenuate ambient noise.

The topics discussed in this thesis include headphone acoustics and measurements, augmented reality audio as well as its applications, and digital filter design. With the help of measurements, the characteristics of the headset could be defined together with its effects on the acoustics of the ear. The measurement results and ARA theory were used to define the required equalization that is needed for acoustical transparency. The equalization was then realized using digital parametric equalizers. Furthermore, since automatic adaptation was one of the target objectives, adaptive filters were used to estimate the isolation online. Finally, the equalization controls were implemented according to the estimated isolation.

The behavior of the proposed algorithm was validated with measurements. The results from these measurements show that the original design objectives were accomplished: the isolation estimation can be performed with ambient sounds, the hear-through equalization results in acoustical transparency, the isolation estimation is utilized to improve the equalization with poor headset fitting, and the algorithm enables user adjustments to enhance the equalization. The adaptive isolation estimation produces similar results to laboratory test performed with controlled signals, and more importantly the results are stable enough to be used as equalization control. In addition, the equalization controls improve the performance of the hear-through equalization: the automatic adaptation improves the resulting responses when compared to constant equalization, the user controls result in noticeable changes that can be used to enhance the hear-through signal, and the comb-filtering effect is avoided with both controls with additional adjustments.

However, there are several aspects regarding the proposed algorithm that could

be addressed in future research. First of all, a mobile implementation should be performed for real-life usability test. The algorithm is intended for everyday use, and therefore its value is estimated in actual use. Since the proposed algorithm forms a digital ARA system together with the prototype headset, a smartphone would be an ideal platform for the algorithm, since no additional hardware is needed apart from the four signal inputs. When implemented on a smartphone, the algorithm could be used anytime, since people carry their phones almost constantly, and thus a comprehensive evaluation could be done. Even though there have been ARA usability test, the adaptability of the proposed algorithm would be a new feature. For example, the speed of adaptability and equalization changes could be analyzed in order to find good values for optimal hear-through experience.

Another future study could focus on the isolation estimation part of the proposed algorithm, which can be detached to a separate entity. It could be utilized, for example, in additional applications, which use the knowledge of whether the headphones are worn or not. However, these applications may need even more precise isolation estimate, and for that altered headset mechanics are needed. With the current prototype headset, the construction and placement of the microphones results in estimation errors, since some frequencies seem to pass through between the microphones. Nevertheless, even with improved construction of the headset earpieces, the microphone placements regulate that the effect of the ear canal cannot be measured, since the internal microphone is at the beginning of the ear canal instead of being near the ear drum, and therefore it must be acknowledged separately.

Furthermore, the isolation properties of the ARA system could be improved for easier control of comb-filtering effect. This can be achieved by utilizing active noise cancellation as a part of the ARA algorithm. An alternative way is to improve the sealing of the headphones, and thus increase the passive attenuation. However, in that case, the coherence between the internal and external microphone signal must be ensured in order to enable the isolation estimation.

In future research, the intelligence of the hear-through function could also be developed. With the proposed algorithm, the user may control the volume of the hear-through signal, which then stays constant until further adjustments. However, the algorithm could be taught to listen for certain important sounds, such as sirens of emergency vehicles that would be let through despite the volume settings. Additionally, such system could be used to improve the safety of pedestrians and cyclists in traffic areas by alarming the user of dangerously incoming vehicles.

References

[1] Gartner, Inc, "Gartner says smartphone sales surpassed one billion units in 2014," http://www.gartner.com/newsroom/id/2996817, press release, Mar. 2015, accessed 24.1.2016.

[2] Gartner, Inc, "Gartner says emerging markets drove worldwide smartphone sales to 19 percent growth in first quarter of 2015," http://www.gartner.com/newsroom/id/3061917, press release, May 2015, accessed 24.1.2016.

[3] ITU-T, "Recommendation P.57, Series P: Telephone transmission quality, objective measuring apparatus—Artificial ears," *International Telecommunication Union, Telecommunication Standardization Sector, Rec. ITU-T P.57*, 2008/1996.

[4] C. A. Poldy, *Headphones,* ch. 14 in Loudspeaker and Headphone Handbook (ed. J. Borwick). Oxford, UK: Focal Press, 3rd ed., 2001.

[5] J. Blauert and N. Xiang, *Acoustics for Engineers: Troy Lectures.* Berlin, Germany: Springer Science & Business Media, 2nd ed., 2009.

[6] G. Wersényi, "On the measurement and evaluation of bass enhanced in-ear phones," in *Proc. 20th International Congress on Acoustics, ICA*, Sydney, Australia, Aug. 2010.

[7] M. Tikander, M. Karjalainen, and V. Riikonen, "An augmented reality audio headset," in *Proc. 11th Int. Conf. on Digital Audio Effects (DAFx-08)*, Espoo, Finland, Sept. 2008.

[8] D. Hammershøi, "Fundamental aspects of the binaural recording and synthesis techniques," in *Proc. AES 100th Convention*, Copenhagen, Denmark, May 1996.

[9] J. Rämö, "Equalization Techniques for Headphone Listening," Ph.D. dissertation, Aalto University School of Electrical Engineering, Dept. of Signal Processing and Acoustics, 2014.

[10] T. D. Rossing, R. F. Wheeler, and P. A. Huber, *The Science of Sound.* Boston, USA: Addison Wesley, 3rd ed., 2009.

[11] R. W. Carlisle, "History and current status of miniature variable-reluctance balanced-armature transducers," *J. Audio Eng. Soc.*, vol. 13, no. 1, pp. 45–49, Jan. 1965.

[12] Etymotic Research, Inc., "ER-4 MicroPro earphones," http://www.etymotic.com/downloads/dl/file/id/72/product/100/er4_micropro_user_manual.pdf, user manual, 2013, accessed 10.11.2015.

[13] N. Kim and J. B. Allen, "Two-port network analysis and modeling of a balanced armature receiver," *Hearing Res.*, vol. 301, pp. 156–167, Jul. 2013.

[14] R. Lichenstein, D. C. Smith, J. L. Ambrose, and L. A. Moody, "Headphone use and pedestrian injury and death in the United States: 2004–2011," *Injury Prevention*, vol. 18, no. 5, pp. 287–290, Oct. 2012.

[15] J. Rämö and V. Välimäki, "Digital augmented reality audio headset," *J. Electr. Comput. Eng.*, vol. 2012, Oct. 2012.

[16] N. Sakamoto, T. Gotoh, and Y. Kimura, "On "out-of-head localization" in headphone listening," *J. Audio Eng. Soc.*, vol. 24, no. 9, pp. 710–716, Nov. 1976.

[17] ITU-R, "Recommendation ITU-R BS.708: Determination of the electro-acoustical properties of studio monitor headphones," *International Telecommunication Union, Radiocommunication Sector, Rec. ITU-R BS.708*, 1997/1990.

[18] G. Lorho, "Subjective evaluation of headphone target frequency responses," in *Proc. AES 126th Convention*, Munich, Germany, May 2009.

[19] S. Olive, T. Welti, and E. McMullin, "Listener preferences for different headphone target response curves," in *Proc. AES 134th Convention*, Rome, Italy, May 2013.

[20] S. Olive, T. Welti, and E. McMullin, "Listener preferences for in-room loudspeaker and headphone target responses," in *Proc. AES 135th Convention*, New York, USA, Oct. 2013.

[21] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858–867, Feb. 1989.

[22] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 868–878, Feb. 1989.

[23] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," *J. Exp. Psychol.*, vol. 27, no. 4, pp. 339–368, Oct. 1940.

[24] O. Kirkeby, "A balanced stereo widening network for headphones," in *Proc. AES 22nd Int. Conf.*, Espoo, Finland, Jun. 2002.

[25] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Design criteria for headphones," *J. Audio Eng. Soc.*, vol. 43, no. 4, pp. 218–232, Apr. 1995.

[26] DIN EN ISO 11904-1, "Determination of sound immission from sound sources placed close to the ear - part 1: Technique using microphone in real ear (MIRE-technique)," *Deutsches Institut für Normung e.V.*, 2003.

[27] M. Hiipakka, M. Tikander, and M. Karjalainen, "Modeling of external ear acoustics for insert headphone usage," *J. Audio Eng. Soc.*, vol. 58, no. 4, pp. 269–281, Apr. 2010.

[28] C. J. Plack, *The Sense of Hearing*. New York, USA: Psychology Press, 2nd ed., 2014.

[29] D. Raichel, *The Science and Applications of Acoustics.* New York, USA: Springer-Verlag, 2000.

[30] T. D. Rossing and N. H. Fletcher *Principles of Vibration and Sound.* New York, USA: Springer-Verlag, 2004.

[31] G. Von Békésy, *Experiments in Hearing*, vol. 8 in Series in Psychology (ed. E. G. Wever). New York, USA: McGraw-Hill, 1960.

[32] L. L. Beranek, *Acoustics.* New York, USA: Acoustical Society of America through the American Institute of Physics, Inc., 1996.

[33] F. M. Wiener and D. A. Ross, "The pressure distribution in the auditory canal in a progressive sound field," *J. Acoust. Soc. Am.*, vol. 18, no. 2, pp. 401–408, Oct. 1946.

[34] P. Brüel, *Impedance of Real and Artificial Ears,* Brüel & Kjær, Application note, 1976.

[35] C. Pörschmann, "One's own voice in auditory virtual environments," *Acta Acust. united Ac.*, vol. 87, no. 3, pp. 378–388, May 2001.

[36] J. Curran, "A forgotten technique for resolving the occlusion effect," *Starkley Audiology Series* reprinted from *Innovations* vol. 2, no. 2, 2012.

[37] F. Kuk, D. Keenan, and C.-C. Lau, "Vent configurations on subjective and objective occlusion effect," *J. Am. Acad. Audiol.*, vol. 16, no. 9, pp. 747–762, Oct. 2005.

[38] J. Kiessling, "Open fittings: Something new or old hat?" in *Hearing Care for Adults First Int. Conf.*, (Chapter 17, pp. 217–226), Stäfa, Switzerland, Mar. 2006.

[39] G. Von Békésy, "The structure of the middle ear and the hearing of one's own voice by bone conduction," *J. Acoust. Soc. Am.*, vol. 21, no. 3, pp. 217–232, May 1948.

[40] A. J. Berkhout, D. de Vries, and M. M. Boone, "A new method to acquire impulse responses in concert halls," *J. Acoust. Soc. Am.*, vol. 68, no. 1, pp. 179–183, Jul. 1980.

[41] A. J. Berkhout, M. M. Boone, and C. Kesselman, "Acoustic impulse response measurement: A new technique," *J. Audio Eng. Soc.*, vol. 32, no. 10, pp. 740–746, Oct. 1984.

[42] P. M. Clarkson, J. Mourjopoulos, and J. K. Hammond, "Spectral, phase, and transient equalization for audio systems," *J. Audio Eng. Soc.*, vol. 33, no. 3, pp. 127–132, Mar. 1985.

[43] D. Griesinger, "Impulse response measurements using all-pass deconvolution," in *Proc. AES 11th Int. Conf.*, Portland, USA, May 1992.

[44] D. Griesinger, "Beyond MLS-Occupied hall measurements with FFT techniques," in *Proc. AES 101st Convention*, Los Angeles, USA, Nov. 1996.

[45] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th Convention*, Paris, France, Feb. 2000.

[46] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. AES 122nd Convention*, Vienna, Austria, May 2007.

[47] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, Jun. 2001.

[48] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262, Apr. 2002.

[49] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *J. Audio Eng. Soc.*, vol. 52, no. 6, pp. 618–639, Jun. 2004.

[50] V. Riikonen, M. Tikander, and M. Karjalainen, "An augmented reality audio mixer and equalizer," in *Proc. AES 124th Convention*, Amsterdam, The Netherlands, May 2008.

[51] M. Tikander, "Usability issues in listening to natural sounds with an augmented reality audio headset," *J. Audio Eng. Soc.*, vol. 57, no. 6, pp. 430–441, Jun. 2009.

[52] P. Hiselius, "Acoustic modelling of earplugs," *Acta Acust. united Ac.*, vol. 92, no. 1, pp. 135–138, Jan. 2006.

[53] D. Brungart, A. Kordik, C. Eades, ad B. Simpson, "The effect of microphone placement on localization accuracy with electronic pass-through earplugs," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, Oct. 2003.

[54] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.*, vol. 100, no. 1, pp. 408–427, Jul. 1996.

[55] T. Lokki, H. Nironen, S. Vesa, L. Savioja, A. Härmä, and M. Karjalainen, "Application scenarios of wearable and mobile augmented reality audio," in *Proc. AES 116th Convention*, Berlin, Germany, May 2004.

[56] M. Karjalainen, M. Tikander, and A. Härmä, "Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[57] C. Dicke, J. Sodnik, and M. Billinghurst, "Spatial auditory interfaces compared to visual interfaces for mobile use in a driving task," in *Proc. Ninth Int. Conf. on Enterprise Information Systems (ICEIS)*, Funchal, Madeira, Portugal, Jun. 2007.

[58] B. F. G. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais, "NAVIG: augmented reality guidance system for the visually impaired," *Virtual Reality*, vol. 16, no. 4, pp. 253–269, Nov. 2012.

[59] R. Väänänen, S. Vesa, and M. Hämäläinen, "Testing the user experience of an augmented reality headset and 3D audio-guided pedestrian navigation" in *Proc. AES 55th Int. Conf.*, Helsinki, Finland, May 2014.

[60] V. Välimäki, A. Franck, J. Rämö, H. Gamper, and L. Savioja, "Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 92–99, Mar. 2015.

[61] J. Rämö and V. Välimäki, "Live sound equalization and attenuation with a headset," in *Proc. AES 51st Int. Conf.*, Helsinki, Finland, Aug. 2013.

[62] S. Brunner, H.-J. Maempel, and S. Weinzierl, "On the audibility of comb-filter distortions," in *Proc. AES 122nd Convention*, Vienna, Austria, May 2007.

[63] A. Clifford and J. D. Reiss, "Using delay estimation to reduce comb filtering of arbitary musical sources," *J. Audio Eng. Soc.*, vol. 61, no. 11, pp. 917–927, Nov. 2013.

[64] B. Rafaely, "Active noise reducing headset—An overview," in *Proc. International Congress and Exhibition on Noise Control Engineering*, The Hague, The Netherlands, Aug. 2001.

[65] W. Ahnert, "Comb-filter distortions and their perception in sound reinforcement systems," in *Proc. AES 84th Convention*, Paris, France, Mar. 1988.

[66] G. Schuller and A. Härmä, "Low delay audio compression using predictive coding," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, Orlando, USA, May 2002.

[67] V. Pulkki and M. Karjalainen, *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. West Sussex, UK: John Wiley & Sons, 1st ed., 2015.

[68] J. O. Smith, *Introduction to Digital Filters with Audio Applications*, `http://ccrma.stanford.edu/~jos/filters/`, online book, 2007 edition, accessed 5.11.2015.

[69] M. Story, "Audio analog-to-digital converters," *J. Audio Eng. Soc.*, vol. 52, no. 3, pp. 145–158, Mar. 2004.

[70] P. F. Titchener, "Adaptive filters for audio—An overview with application examples," in *Proc. AES 93rd Convention*, San Francisco, USA, Oct. 1992.

[71] B. Baker, "Analogue-to-digital conversion: Basics of ADC latency," *EE Times India*, Mar. 2009.

[72] R. Yates, and R. Lyons, "DC blocker algorithms," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 132–134, Mar. 2008.

[73] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing.* New York, USA: John Wiley & Sons, 1st ed., 1993.

[74] K. Steiglitz, *A Digital Signal Processing Primer with Applications to Digital Audio and Computer Music.* Menlo Park, USA: Addison Wesley, 1st ed., 1996.

[75] R. Berkovitz, "Digital equalization of audio signals," in *Proc. AES 1st Int. Conf.*, Rye, USA, Jun. 1982.

[76] A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.*, vol. 60, no. 1/2, pp. 54–62, Jan./Feb. 2012.

[77] S. E. Olive and T. Welti, "The relationship between perception and measurement of headphone sound quality," in *Proc. AES 133rd Convention*, San Francisco, USA, Oct. 2012.

[78] P. F. Hoffmann, F. Christensen, and D. Hammershøi, "Insert earphone calibration for hear-through options," in *Proc. AES 51st Int. Conf.*, Helsinki, Finland, Aug. 2013.

[79] D. A. Bohn, "Operator adjustable equalizers: An overview," in *Proc. AES 6th Int. Conf.*, Nashville, USA, May 1988.

[80] P. Regalia and S. Mitra, "Tunable digital frequency response equalization filters," *IEEE Trans. Acoust. Speech*, vol. 35, no. 1, pp. 118–120, Jan. 1987.

[81] D. A. Bohn, "Constant-Q graphic equalizers," *J. Audio Eng. Soc.*, vol. 34, no. 9, pp. 611–626, Sept. 1986.

[82] U. Zölzer and T. Boltze, "Parametric digital filter structures," in *Proc. AES 99th Convention*, New York, USA, Oct. 1995.

[83] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation.* New York, USA: Springer, 3rd ed., 2008.

[84] S. Haykin, *Adaptive Filter Theory* in Prentice Hall Information and System Sciences Series (ed. T. Kailath). New Jersey, USA: Prentice Hall, 4th ed., 2002.

[85] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *WESCON Conv. Rec.,* pt. 4, pp. 96–104, 1960.

[86] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* in Prentice Hall Signal Processing Series (ed. A. V. Oppenheim). New Jersey, USA: Prentice Hall, 1985.

[87] J. I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Control*, vol. 12, no. 3, pp. 282–287, Jun. 1967.

[88] L. R. Vega, H. Rey, J. Benesty, and S. Tressent, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.

[89] N. A. Verhoeckx, H. C. Van Den Elzen, F. A. Snijders, and P. J. Van Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 6, pp. 768–781, Dec. 1979.

[90] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, vol. 8 of Communication and Cybernetics. New York, USA: Springer-Verlag, 1976.

[91] G. U. Yule, "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers," *Phil. Trans. R. Soc. Lond. A*, vol. 226, pp. 267–298, Apr. 1927.

[92] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.*, vol. 50, no. 2, pp. 637–655, Aug. 1971.

[93] S. H. Choi, K. K. Kim, and H. S. Lee, "Speech recognition using quantized LSP parameters and their transformations in digital communication," *Speech Commun.*, vol. 30, no. 4, pp. 223–233, Apr. 2000.

[94] A. Härmä and U. K. Laine, "A comparison of warped and conventional linear predictive coding," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 579–588, Jul. 2001.

[95] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[96] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, Nov. 2000.

[97] A. G. Constantinides, "Spectral transformations for digital filters," *Proc. IEE*, vol. 117, no. 8, pp. 1585–1590, Aug. 1970.

[98] A. V. Oppenheim, D. H. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast Fourier transform," *Proc. IEEE*, vol. 59, no. 2, pp. 299–301, Feb. 1971.

[99] A. V. Oppenheim and D. H. Johnson, "Discrete representation of signals," *Proc. IEEE*, vol. 60, no. 6, pp. 681–691, Jun. 1972.

[100] C. Braccini and A. V. Oppenheim, "Unequal bandwidth spectral analysis using digital frequency warping," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-22, no. 4, pp. 236–244, Aug. 1974.

[101] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Am.*, vol. 68, no. 4, pp. 1071–1076, Oct. 1980.

[102] J. Huopaniemi, "Virtual Acoustics and 3-D Sound in Multimedia Signal Processing," Ph.D. dissertation, Helsinki University of technology, Laboratory of Acoustics and Audio Signal Processing, 1999.

[103] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.*, vol. 47, no. 1-2, pp. 103–138, Aug. 1990.

[104] B. C. Moore, R. W. Peters, and B. R. Glasberg, "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 132–140, Jul. 1990.

[105] P. F. Hoffmann, F. Christensen, and D. Hammershøi, "Insert earphone calibration for hear-through options," in *Proc. AES 51st Int. Conf.*, Helsinki, Finland, Aug. 2013.

[106] R. Humphrey, "Playrec, multi-channel Matlab audio" `http://www.playrec.co.uk/`, 2008–2014, accessed 11.1.2016.

[107] Head Acoustics GmbH, "Artificial Head System HMS II.3," `https://www.head-acoustics.de/downloads/eng/hms_II/D1230_HMS_II_3_e.pdf`, data sheet, 2015, accessed 23.1.2016.

[108] MOTU, Inc., "UltraLite-mk3 Hybrid User Guide for Mac," `http://cdn-data.motu.com/manuals/firewire-usb-audio/UltraLite-mk3_Hybrid_Mac.pdf`, user guide, 2010, accessed 23.1.2016.

[109] Genelec Oy, "Genelec 8030B Active Monitoring System," `http://www.genelec.com/sites/default/files/media/Studiomonitors/8000SeriesStudioMonitors/8030B/genelec_8030b_datasheet.pdf`, data sheet, 2013, accessed 23.1.2016.